



INSTITUTO TECNOLÓGICO SUPERIOR DE MISANTLA

DEPARTAMENTO DE COMPUTACIÓN

REDES GENERATIVAS ADVERSARIAS EN
IMÁGENES DE NANOESTRUCTURAS DE TCO

T E S I S

PARA OBTENER EL GRADO DE:
Maestro en Sistemas Computacionales

PRESENTA:

I.S.C. Karim González Tovar

DIRECTOR:

Dr. Rajesh Roshan Biswal

CO-ASESOR:

Dr. Eddy Sánchez de la Cruz

Misantla, Veracruz, México, Agosto 2018

Dedico el presente trabajo a todas las personas que me apoyaron y creyeron en mí, en especial a:

A mi madre Martha Patricia

Por la tenacidad y constancia que siempre ha demostrado, motivándonos a la preparación profesional y motivación constante que me ha permitido ser una persona de bien, pero más que nada, por su amor y paciencia.

Gracias

A mi abuelita Martha

Que siempre cuidó a sus nietos de la mejor manera, inculcando valores de educación y respeto.

Que en paz descanse, gracias

A mis hermanas

A mis hermanas Karen y Grecia por estar ahí molestando.

A mis familiares

A mis tíos, tías, primos y primas que cada vez somos más unidos.

Agradecimientos

Quiero agradecer primeramente a mis amigos, conocidos y compañeros de maestría por su gran amistad y ayuda incondicional. A los docentes con los que tuve el honor de tomar clases y en especial; a Dr. Luis Alberto Morales Rosales por aceptarme en el programa de maestría en sistemas computacionales, a el director de tesis Dr. Rajesh Roshan Biswal, a Dr. Héctor Rodríguez Rangel por sus clases y amistad así como también a Dr. Simón Pedro Arguijo Hernández por sus sabios concejos y magistrales clases, Dr. Eddy Sánchez de la Cruz. También quisiera agradecer al Concejo Nacional de Ciencia y Tecnología *CONACyT* por la beca de manutención de periodo Febrero 2016 - Enero 2018 con CVU 742177, sin este apoyo el presente trabajo de investigación se podría realizar.

Resumen

In this work, we propose Conditional Generative Adversarial networks (CGAN) for the deposition of TCO thin films in microscopical SEM images. In order to investigate morphological characterization of TCO films in nanomaterials and discover new appealing applications, the following research is highly useful. We used an extended form of GAN called cGANs to perform simulations of the morphological results using the SEM images of ZnO thin films, considering a few image acquisition parameters. Modeling using semi supervised deep learning approach have shown incredible results in problems related to computer vision such as image translation, image generation, image synthesis, simulation etc. We demonstrated that training a Deep Generative model, can be used to obtain good result by generating images artificially, good images take in consideration sparse raw images. Also, the generated images were validated with quantitative metrics by inception score.

Índice general

Índice de figuras	ix
Índice de tablas	x
1. Generalidades	1
1. Introducción	1
2. Planteamiento del problema	2
3. Justificación	3
4. Hipótesis	4
5. Objetivos	5
5.1. Objetivo general	5
5.2. Objetivos específicos	5
6. Propuesta de solución	6
7. Alcances y limitaciones	8
8. Estructura de la tesis	8
2. Marco teórico	10
1. Nanotecnología	10
1.1. Nanomateriales	11
1.2. Óxidos transparentes conductivos	11
1.3. Película delgada transparente	12
1.4. Proceso de creación de películas delgadas	13
1.5. Métodos de deposición	14
1.6. Caracterización	14
1.7. Microscopía	17
2. Ciencias de la computación	18
2.1. Inteligencia artificial	18
2.2. Aprendizaje Automático	19
2.3. Aprendizaje supervisado	20
2.4. Aprendizaje no supervisado	20

2.5.	Aprendizaje semi-supervisado	20
2.6.	Aprendizaje profundo	21
2.7.	Redes convolucionales	21
2.8.	Deconvolución	22
2.9.	Modelos generativos	22
2.10.	Auto-codificadores	23
2.11.	Auto-codificadores variacionales	23
2.12.	Redes generativas adversarias	23
2.13.	Función costo	25
2.14.	Comparación de modelos generativos	26
2.15.	Redes generativas adversarias condicionales	27
2.16.	Arquitectura de las redes generativas	28
3.	Métricas de evaluación de modelos	30
3.1.	Trabajos relacionados	30
3.	Metodología	33
1.	Descripción de la metodología	33
1.1.	Adquisición de información	34
2.	Descripción de las imágenes	34
3.	Aumentación de datos y pre-procesamiento	35
3.1.	Selección de imágenes	35
3.2.	Pre-procesado de imágenes	36
4.	Aumentación de datos	38
5.	Modelado	40
5.1.	Modelo	40
5.2.	Arquitectura	41
5.3.	Entrenando GANs	43
5.4.	Programación	44
5.5.	Evaluación y pruebas	44
4.	Resultados	45
1.	Experimentos	46
1.1.	Resultados cualitativos	48
2.	Generación condicionada	50
3.	Resultados cuantitativos	52
3.1.	Análisis de las tablas	52
3.2.	Análisis del entrenamiento	53
5.	Conclusiones y trabajo futuro	54
1.	Conclusiones	54
2.	Trabajo futuro	55

A. Código/Librerías	56
1. Código	56
2. Librerías	56
3. Lista de modelos pre-entrenados	57
Bibliografía	58

Índice de figuras

1.1. Esquema general de arquitectura	6
1.2. Diagrama de flujo pre-proceso y modelado	7
2.1. Métodos de deposición	14
2.2. Esquema de flujo de caracterización	15
2.3. Diagrama de flujo de caracterización	16
2.4. Operación de convolución	22
2.5. Diagrama de sub área en GAN	28
3.1. Muestra de imágenes (pre-procesadas) del dataset	35
3.2. Imagen con leyendas	36
3.3. Acercamiento a las imágenes	37
3.4. Resultados con diferentes filtros	38
3.5. Cálculo de filtros	39
3.6. Cálculo de filtros	40
3.7. Diagrama en bloques de las redes generador G y discriminador D	42
4.1. Muestra de imágenes que no convergen	47
4.2. Muestras de generación con VAE	48
4.3. Visualización de entrenamiento	49
4.4. Generaciones condicionadas I	51
4.5. Generaciones condicionadas II	51
4.6. Progreso de entrenamiento con 120 épocas	53

Índice de tablas

3.1. Tabla de librerías	44
4.1. Tabla generación condicionada	52

Capítulo 1

Generalidades

1 Introducción

La nanotecnología y nano-materiales se considera el campo más innovador del siglo *XXI*, las investigaciones llevadas a cabo en la fabricación de nuevos materiales son extensas y tienen aplicabilidad directa a productos cotidianos para cualquier persona, aunque la nanotecnología se refiere el estudio de objetos con longitud de 1×10^{-9} , sus aplicaciones la podemos encontrar en cremas, sensores, celulares y bastantes más. Algunos avances incluyen la impregnación de sustancias u objetos adherentes para darles ciertas propiedades adicionales, que viene siendo una aplicación de los óxidos transparentes conductivos en forma de películas delgadas.

Las películas delgadas son un material de óxido de zinc **ZnO** depositado en películas delgadas tiene una gran variedad de propiedades físicas controladas tales como conductividad, reflectancia, transparencia en el espectro visible que pueden ser usadas en la industria de manufactura de dispositivos electrónicos y opto electrónicos.

Cuando la película delgada ha sido depositada los humanos no podemos ver algo tan pequeño como una nano-estructura, por lo que la ciencia encargada de analizar y medir objetos microscópicos es la optometría. La optometría se encarga también del proceso de adquirir imágenes digitales y también procesar imágenes para un análisis más conveniente. El proceso científico en nano-materiales es una tarea cien por ciento experimental, que, el problema de experimentación es que, para poder observar nano-estructuras fabricadas por condiciones específicas, experimentación no se tienen modelos computacionales precisión capaz de predecir la morfología del óxido de zinc depositado en forma de película delgada, se tiene que realizar paso a paso la experimentación para analizar resultados.

Los métodos empleados para deposición de películas delgadas pueden ser divididos dentro de dos grupos físicos y químicos. Los métodos físicos incluyen depósito de vapor físico, ablación láser, haz molecular y sputtering. Los métodos químicos comprenden métodos de deposición fase-gaseosa y técnicas de solución. Los métodos de fase gaseosa son deposición química a base de vapor. Los métodos que emplean soluciones precursoras son rocío pirolítico, sol-gel y baño químico.

Los desarrollos de nuevas tecnologías siempre inculcan áreas del conocimiento multidisciplinario en cualquier área de las ingenierías, en este caso la revolución tecnología y la nanotecnología y la versatilidad de las ciencias de la computación. En el área de computación han sido desarrollados recientemente algoritmos, métodos y arquitecturas para la generación de imágenes, a este conjunto se le conoce como aprendizaje profundo (en inglés, Deep learning) que son un conjunto de modelos y técnicas para resolver problemas complejos utilizando redes neuronales artificiales, arquitectura y operaciones especiales. Con un modelo generativo se puede abstraer el proceso de experimentación en la fabricación de películas delgadas de ZnO mediante rocío químico pirolítico, asimismo visualizar con imágenes realistas la salida del modelo.

2 Planteamiento del problema

Dentro del área de materiales se encuentra la sub-área de nanomateriales que han tenido grandes avances tecnológicos en años recientes, en parte a los nuevos desarrollos en microscopía e investigaciones en química, física, biología, computación. Mientras esto ocurría las aplicaciones en nanotecnología fueron creciendo en muchas áreas como, por ejemplo; en medicina, materiales, física, robótica, etc.

Actualmente la investigación en nanomateriales es costosa porque se requieren recursos materiales para realizar experimentos en equipos especializados. Para poder investigar materiales en la forma de película delgada se necesita el desarrollo de nuevas tecnologías para poder entender las propiedades físicas de los materiales en esta forma según la aplicación.

En los últimos años se han venido realizado grandes esfuerzos para tratar de entender y poder predecir las propiedades físicas de dichos materiales fabricados en forma de película delgada. El avance ha sido considerable, al grado de poder explotarse los conocimientos adquiridos en la elaboración de dispositivos optoelectrónicos y en diversas aplicaciones tecnológicas.

El depósito de estos materiales en un área grande es otra parte importante para la producción de los dispositivos con métodos que deben tener bajos costos. Los métodos que reúnen estos requisitos son el de solución química que incluye sol-gel, electro-depósito y rocío químico, que son técnicas viables para la producción de dispositivos foto-voltaicos comerciales. Rocío químico pirolítico a diferencia de las otras técnicas de deposición de películas, representa un proceso muy simple y relativamente de bajo costo. Ofrece una técnica sumamente fácil para preparar películas de cualquier composición incluyendo ZnO. Además, no requiere sustratos o químicos de alta calidad.

3 Justificación

En México la investigación experimental en nanotecnología está reservada a pocas instituciones con recursos suficientes para poseer equipo especializado, lo que convierte a dicha área muy difícil para realizar investigación y/o producción de películas delgadas (en inglés, thin films).

El óxido de zinc se encuentra abundante en el planeta y es barato comparado con otros óxidos semiconductores. El óxido de zinc está siendo utilizado en manufactura comercial fabricando televisiones de cristal líquido, paneles de plasma, televisiones OLED. En forma de película delgada las podemos encontrar en dispositivos opto electrónicos como; televisiones, paneles solares, teléfonos celulares componentes electrónicos, ventanas inteligentes. Todo esto a gracias a sus propiedades físicas como; estabilidad química, transparencia en el espectro visible y conductividad.

Los recursos tecnológicos que refieren al depósito de películas delgadas son bastante costosos, pocas instituciones del país y de latino-américa cuentan con novedosos equipos para fabricar películas delgadas, también se requiere equipo para analizar las muestras obtenidas con equipos de microscopía y sus diferentes técnicas, microscopía electrónica de transmisión TEM, microscopía electrónica de barrido SEM, microscopía electrónica de transmisión de barrido SEM y microscopía de fuerza atómica AFM.

La experimentación sucede a prueba y error para obtener y describir resultados, que se tienen que repetir una y otra vez con diferentes condiciones de fabricación por lo que pueden llegar a ser tardados y costosos. En la figura 2.2 se muestra la representación del proceso.

A diferencia de muchas otras técnicas de deposición de películas, rocío pirolítico representa un proceso muy simple y relativamente de bajo costo. Ofrece una técnica sumamente fácil para preparar películas de cualquier composición. Además, no requiere sustratos o químicos de alta calidad.

Aunque el ZnO obtenidos por métodos físicos muestran la mejor característica debe mencionarse que bajo ciertas condiciones controladas es posible obtener también películas de calidad basadas en técnicas químicas menos costosas. Otra ventaja sumamente importante de la técnica de rocío pirolítico es que uno puede variar fácilmente la concentración del dopante y por lo tanto variar las propiedades de las películas.

El tiempo que se requiere para realizar el análisis de nano-estructuras es demasiado largo y depende de la técnica de deposición, así recorriendo el siguiente flujo: preparación y calibración de equipo, deposición de material, proceso de fabricación de películas delgadas, microscopía y por ultima guardar resultados y etiquetarlos de forma digital.

4 Hipótesis

Es posible generar simulaciones de imágenes de microscopía SEM de nano-estructuras en imágenes de películas delgadas de óxido de zinc depositadas con la técnica de rocío químico pirolítico utilizando parámetros de obtención de la imagen .

5 Objetivos

5.1 Objetivo general

Entrenar un modelo computacional de aprendizaje profundo que genere imágenes de nano-estructuras de óxido de zinc en base a parámetros de obtención.

5.2 Objetivos específicos

Objetivos específicos son presentados en orden de desarrollo.

1. Selección de imágenes.
2. Pre-procesar imágenes para quitar ruido y estandarizar entradas al modelo.
3. Utilizar mecanismos de aumentación de datos.
4. Construir la base de conocimientos a partir de las imágenes y sus datos.
5. Implementar arquitectura generacional de aprendizaje profundo.
6. Entrenar red neuronal generativa adversaria.
7. Análisis de resultados del modelo generativo.

6 Propuesta de solución

La propuesta de solución incluye diseñar un modelo computacional basado en redes neuronales profundas con una arquitectura llamada redes neuronales generativas adversarias, por medio de este modelo generar pequeñas imágenes de nano-estructuras que simulan la digitalización de ciertos parámetros. Se basa en la sub-área de aprendizaje automático aprendizaje semi-supervisado utilizando técnicas y modelos del aprendizaje profundo. El presente trabajo se centra en entrenar un modelo computacional, donde las entradas del modelo son imágenes microscopias adquiridas en procesos experimentales y sus parámetros de obtención. La metodología consiste en construir una base de conocimientos con las imágenes, pre-procesar las imágenes, aumentar el dataset, entrenar el modelo y validar. Figura 2.2

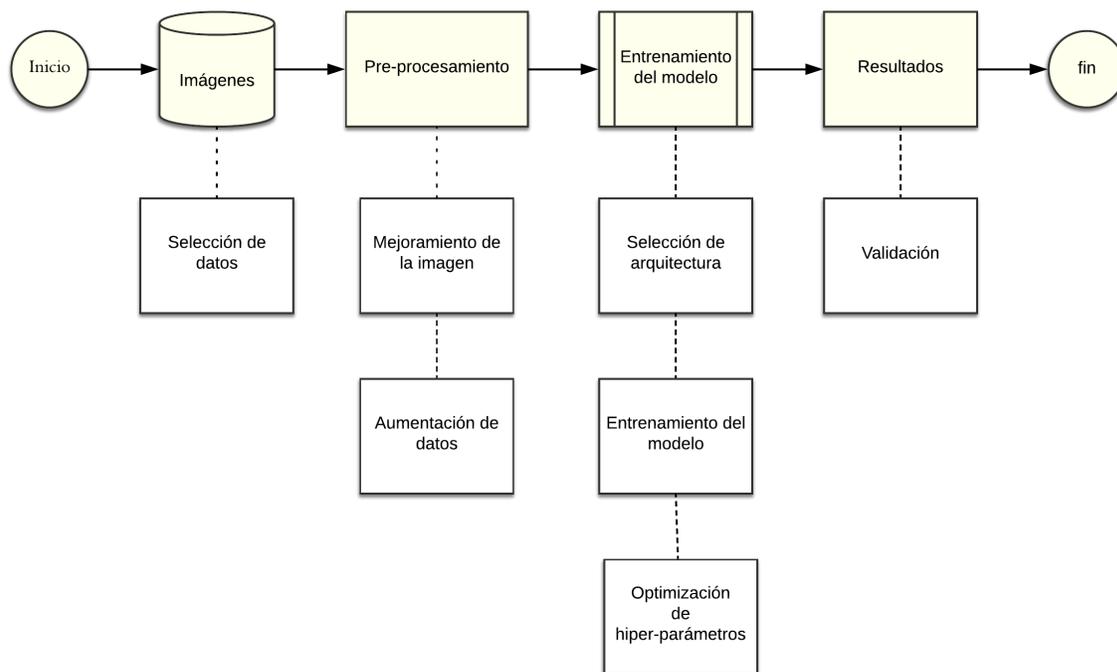


Figura 1.1: Esquema general de arquitectura

Se necesita pre-procesar las imágenes para normalizar su resolución, para después en base a la arquitectura de redes GAN, construir la arquitectura de las redes neuronales generativo y discriminador, así como analizar y seleccionar los hiper-parámetros más adecuados. Para finalizar evaluar el rendimiento y precisión del modelo con métricas específicas y usando parámetros de obtención de películas delgadas, cabe resaltar que las imágenes nunca son vistas por el modelo generador en la etapa de entrenamiento, la salida será una imagen generada por el modelo que representa la nano-estructura depositada en el sustrato con dichos parámetros.

En la siguiente figura se 1.2 muestran de manera general secciones de pre-procesamiento y modelado. Pre-procesado y selección necesarios para un mejor desempeño, como resultado un nuevo dataset más limpio. El modelo generador tiene como entradas un vector aleatorio z concatenado a variables condicionadas c , el discriminador tiene como entradas el mismo vector de variables condicionadas c , imágenes intercaladas de imágenes generadas por el generador G e imágenes reales, 0 si clasifica la muestra proviene de G como falsa, 1 si la muestra es real.

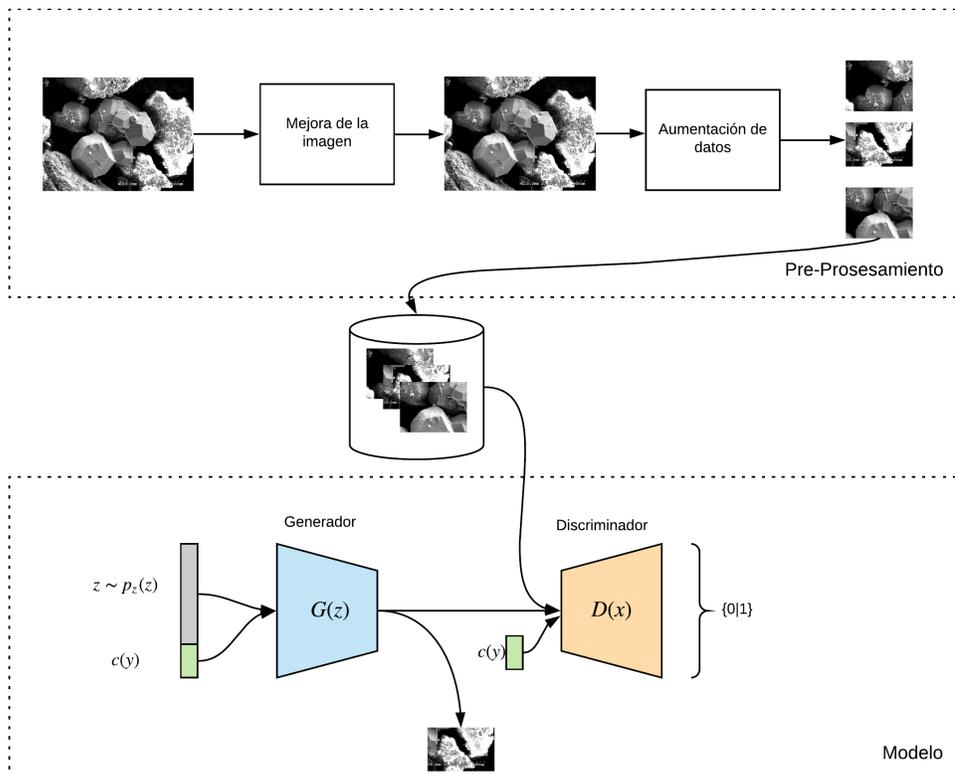


Figura 1.2: Diagrama de flujo pre-proceso y modelado

La evaluación y pruebas son utilizando métricas especializadas para este tipo de modelos *inception score*, sin embargo, el criterio de vista también es importante al verificar que las imágenes realmente se parezcan a las contenidas en el dataset.

7 Alcances y limitaciones

La variedad de óxidos transparentes conductivos es relativamente poca: TCO basados en Selenio, Zinc, Indio, Magnesio, Titanio, Estaño, Germanio, Galio. Cada material con procesos de fabricación de nano-estructuras que pueden ser diferentes, parámetros de fabricación diferentes y sobre todo nano-estructuras diferentes. Por lo que este trabajo se centra en la generación de nano-estructuras de óxidos transparentes conductivos en forma de película delgada basados en óxido de zinc exclusivamente.

En cuanto a las condiciones de creación, la abstracción del proceso de fabricación a trabajar es una llamada rocío pirolítico ultrasónico, aunque existen otras como por ejemplo: sputering, sol-gel, láser pulsado en figura (2.1). El proceso de deposición de las películas delgadas fue rocío pirolítico ultrasónico, posteriormente la digitalización mediante equipo de microscopia SEM. Los parámetros de simulación se obtuvieron de las leyendas de cada imagen (ver figura 3.2), donde se pueden apreciar distancia, magnificación, energía. Esta información se utiliza en el modelo de proceso de simulación. Ver figura 1.2.

Los modelos generacionales que fueron tomados en cuenta a base de tecnología en estado de arte de aprendizaje profundo, las cuales son; redes generativas adversarias, redes codificadoras-decodificadoras, redes codificadoras variacionales, son las que nos permiten generar imágenes de mejor calidad y en caso de las condicionales nos permiten condicional la generación, útil para objetivos de investigación.

8 Estructura de la tesis

El presente trabajo está dividido en cinco capítulos. Al principio se encuentra la sección de generalidades donde se describen. Capítulo 2 describe la base teórica el trabajo, está dividida en dos secciones la primera cubre definiciones de el área de nanotecnología y y la segunda definiciones en computación desde redes neuronales artificiales hasta modelos generativos. Capítulo 3 describe la metodología que fue utilizada para la realización de el presente trabajo, Capítulo 4 análisis cualitativo y cuantitativo de imágenes generadas y también gráficas de resultados, Capítulo 5 conclusiones de los

resultados y propuestas sobre trabajo futuro que incluye ideas y posibles mejoras. Para finalizar la sección apéndice [A](#) que contiene hyper-vínculos de código desarrollado además librerías y fuentes web.

Capítulo 2

Marco teórico

1 Nanotecnología

El termino “nano” proviene del griego que significa pequeño, es una medida de longitud que comprende el rango de 1×10^{-10} metros, realmente muy pequeño para que un ser humano lo divise, para poner en contexto visual, un humano puede ver 10^{-2} metros y una molécula de una molécula de ADN es dos nanómetros. Objetos “nano” como las películas delgadas tienen la particularidad de la relación que tienen entre volumen y superficie es mucho mayor es decir que fórmula magnifican sus propiedades físicas de conductividad, transparencia.

La palabra nanotecnología se usa extensivamente para definir las ciencias y técnicas que se aplican al nivel de nanoescala, esto es, medidas extremadamente pequeñas, que permiten trabajar y manipular las estructuras moleculares y sus átomos. En síntesis conduciría a la posibilidad de fabricar materiales y máquinas a partir del reordenamiento de átomos y moléculas. La nanotecnología se define como el estudio, diseño, creación, síntesis, manipulación y aplicación de materiales, aparatos y sistemas funcionales a través del control de la materia a nanoescala, y la explotación de fenómenos y propiedades de la materia a nanoescala.

Cuando se manipula la materia a la escala tan minúscula de átomos y moléculas, demuestra fenómenos y propiedades totalmente nuevas. Por lo tanto, científicos utilizan la nanotecnología para crear materiales, aparatos y sistemas novedosos y poco costosos con propiedades únicas. La revolución nanotecnológica, se asocia, por una parte, a la fabricación molecular cuya viabilidad tendría un impacto enorme en nuestras vidas, en las economías, los países y en la sociedad en general en un futuro no lejano. Entre los efectos, destacan su potencial impacto en la medicina, la biología, el medio ambiente,

la informática, la construcción. En la actualidad los principales avances prácticos ya se dan en algunos campos: nanopartículas, nanotubos. Los progresos en materia de nanorobots y auto-reproducción son objeto de polémica entre los expertos. Lo que no cabe duda es que la revolución ha comenzado. Y también el debate sobre sus beneficios y riesgos.

1.1 Nanomateriales

Los nanomateriales son materiales con propiedades morfológicas más pequeñas que un micrómetro en al menos una dimensión. A pesar del hecho de que no hay consenso sobre el tamaño mínimo o máximo de un nanomaterial, algunos autores restringen su tamaño de 1 a 100 *nm*, una definición lógica situaría la nanoescala entre la microescala (1 micrómetro) y la escala atómica molecular (alrededor de 0.2 nanómetros).

El ZnO es un semiconductor del grupo II-VI del tipo n. La adición de impurezas para modular sus propiedades eléctricas (doping) nativa del semiconductor debida a las vacantes del oxígeno o intersticiales de zinc es tipo *n*.

Este semiconductor tiene diversas propiedades favorables, incluyendo buena transparencia, alta movilidad de electrones, amplio rango de energía donde no existen estados electrónicos, y fuerte luminiscencia a temperatura ambiente. Estas propiedades son importantes para las aplicaciones emergentes: electrodos transparentes en aparatos de cristal líquido, ventanas ahorradoras de energía o protectoras del calor, y electrónicos como transistores de película delgada y diodos emisores de luz.

1.2 Óxidos transparentes conductivos

Un Óxido Conductor Transparente (*TCO por sus siglas en inglés*) es un semiconductor de banda prohibida ancha, y como su nombre lo indica, tiene una relativa transparencia inherente en la región visible y es altamente conductor debido a la presencia de una alta concentración de electrones libres en su banda de conducción. Esto es debido a defectos en el material o por dopantes extrínsecos, es decir, niveles de impureza cercanos al borde de la banda de conducción. Un TCO es un compromiso entre la conductividad eléctrica y transmitancia óptica.

El primer reporte de un *TCO* fue publicado en 1907 cuando Badeker reportó que una película delgada de cadmio metálico podría ser transparente si se oxidaba, sin perder su propiedad de conducir electricidad (Badeker, 1907). Desde entonces, el valor comercial de estas películas delgadas ha sido reconocido, y la lista de posibles TCOs

se amplía para incluir, por ejemplo, In_2O_3 dopado con Sn (Indium Tin Oxide: ITO), SnO_2 dopado con F (FTO), SnO_2 dopado con Sb (ATO), ZnO dopado con Al (AZO), ZnO dopado con In (IZO), entre otros. Desde la década de 1960, el TCO más ampliamente utilizado para aplicaciones de dispositivos optoelectrónicos ha sido el ITO.

Actualmente, y probablemente en el futuro, este material ofrece el mejor rendimiento disponible en términos de conductividad y transmisión óptica, combinada con una excelente estabilidad del medio ambiente, la reproducibilidad y buena morfología de la superficie. El uso de estos materiales ha sido como conductores eléctricos transparentes para una amplia variedad de aplicaciones en la optoelectrónica, a saber: electrodos transparentes para pantallas de cristal líquido (LCDs), películas delgadas transparentes, diodos emisores de luz, sensores-UV y celdas solares (Ohta H. y Hosono, 2004)

1.3 Película delgada transparente

Una película conductora transparente (en inglés, Transparent Conducting Films TCO) es una capa delgada ópticamente transparente y eléctricamente conductora. Las TCO para aplicaciones fotovoltaicas se han fabricado a partir tanto de materiales inorgánicos, como de orgánicos. Las películas inorgánicas típicamente se componen de una capa de óxido conductor transparente (TCO), generalmente en forma de óxido de indio y estaño (ITO), óxido de estaño dopado con flúor (FTO) y óxido de zinc dopado.

Las películas conductoras transparentes actúan como una ventana para que la luz pase a través del material activo por debajo (donde se produce la generación de portador), como un contacto óhmico para transporte de portadores de la fotovoltaica y también puede actuar como portador transparente para dispositivos de montaje en superficie utilizados entre el cristal laminado o compuestos transmisores de luz. Los materiales transparentes poseen bandas prohibidas con energías que corresponden a longitudes de onda que son más cortas que el intervalo visible de 380 nm a 750 nm . Como tal, los fotones con energías por debajo de la banda prohibida no se recogen por estos materiales y, por lo tanto, la luz visible pasa a través. Sin embargo, las aplicaciones tales como la energía fotovoltaica pueden requerir un intervalo de banda incluso más amplio para evitar la absorción no deseada de los espectros de energía solar.

Los óxidos conductores transparentes (TCO) son de gran importancia tecnológica. Por ejemplo, son usados como: i) electrodos transparentes, ii) escudo de interferencia electromagnética antiestática y iii) como un calentador eléctrico, entre otras. Actualmente se han alcanzado resistividades del orden de $10^{-4} \omega \text{-cm}$ en películas delga-

das(PDs) de ZnO impurificadas con In ó Al, usando la técnica de erosión catódica. El ITO es el TCO más utilizado hoy en día debido a su combinación de la resistividad de cuadro y su transparencia óptica . Aún cuando su producción supera con mucho a la de otros TCO su costo es cada vez más salto debido al encarecimiento del In por ser un elemento poco abundante en la naturaleza. Por lo anterior, es importante el estudio de otros TCO en PDs utilizando técnicas de preparación económicas y sencillas de implementar; como es la técnica de sol-gel.

Por su sus propiedades ópticas, el ZnO es un buen candidato a substituir al ITO , sin embargo sus valores de resistividades son relativamente altas cuando se prepara por sol-gel. Por otro parte el CdO presenta resistencias de cuadro del orden pero su ventana óptica es menor a la del ZnO . En estudios previos se ha encontrado que al combinar ambos óxidos sus propiedades ópticas y eléctricas se combinan y podrían dar lugar a un TCO de buena calidad; similar a la del ITO .

1.4 Proceso de creación de películas delgadas

En la actualidad para poder investigar los materiales en forma de película delgada se necesita del desarrollo de nuevas tecnologías, así como la creación de materiales, sistemas y dispositivos, que nos permitan entender las propiedades físicas de los materiales según su aplicación y requerimiento de acuerdo al avance del mundo moderno. La meta primordial en la física del estado sólido es descubrir materiales novedosos y darles una aplicación, es decir, incorporarlos en diferentes áreas tales como la biofísica, optoelectrónica, nanotecnología entre otros; para lograrlo, es importante conocer los procesos fundamentales y mecanismos microscópicos que rigen el depósito de películas, esto con el fin de adecuarlos a los requerimientos que se exigen ([Enrique Josué Chan Díaz et al., 2012](#)).

Varias técnicas de proceso de deposición en crecimiento de la película se han desarrollado en años recientes para la síntesis de óxido de zinc; descomposición termal, síntesis hidro-termal, sol-gel. Además, las películas delgadas pueden ser depositadas por métodos químicos; deposición de vapor química. la caracterización de películas delgadas como un proceso de análisis, las películas delgadas son sujetas a la caracterización de propiedades morfológica, estructural y óptica. En donde cada propiedad es cuantificada.

1.5 Métodos de deposición

Existen varias técnicas de crecimiento (Biswal et al., 2013), incluyendo RF/DC sputtering, deposición de vapor químico, sol-gel, laser pulsado, rocío pirolítico an sido empleados para obtener películas delgadas de ZnO sobre varios sustratos , el método de rocío pirolítico es considerado como una opción no costosa, simple y flexible para sintetizar áreas grandes de cualquier material. Dicho proceso permite un buen control sobre material dopante. Además se requiere baja temperatura para la cristalización. Además cuenta con las siguientes ventajas; bajo coste de fabricación , Bajo coste para la investigación de estructura y propiedades ópticas.

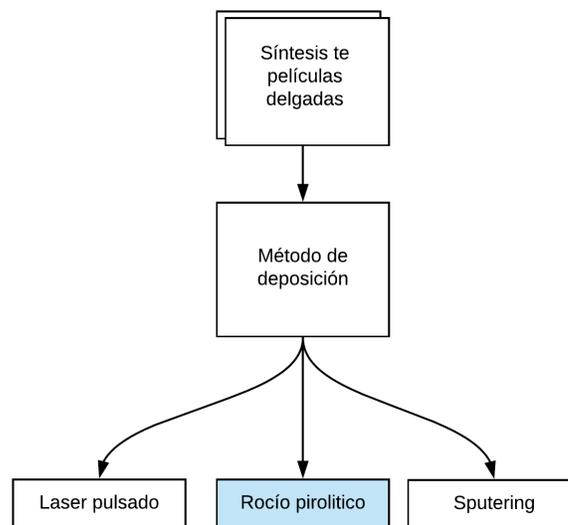


Figura 2.1: Métodos de deposición

1.6 Caracterización

Propiedades físicas de los TCOs los parámetros físicos de deposición tienen una influencia directa en las propiedades morfológicas de nano-estructuras, las propiedades morfológicas de las películas delgadas pueden ser analizadas usando microscopía usando la técnica SEM, las propiedades morfológicas de las películas delgadas están fuertemente relacionadas con el óxido a depositar, técnica de deposición, dopado del óxido entre otros aspectos, ver figura 2.2.

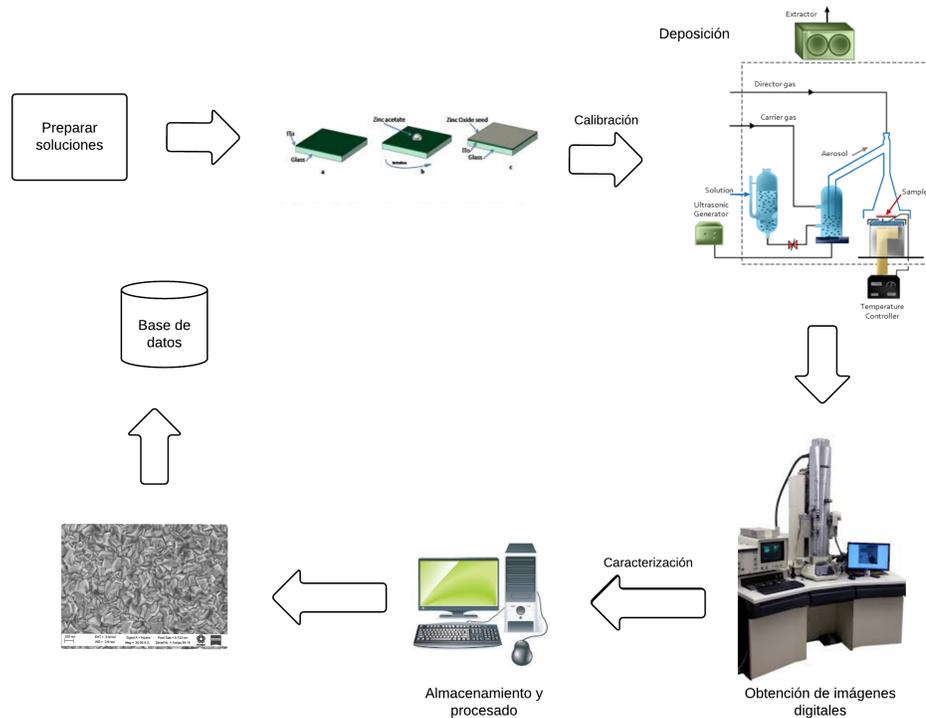


Figura 2.2: Esquema de flujo de caracterización

Morfología el óxido de zinc se cristaliza en dos formas principalmente; en hexágonos y cubos. La estructura hexagonal es más estable en condiciones normales. La forma de blenda de zinc puede ser estabilizada al formar el ZnO en sustratos con estructura cúbica enrejada. En ambos casos, el centro del cinc y el óxido son tetraédricos, la más característica geometría del $Zn(II)$. El ZnO se convierte en sal de roca a presiones relativamente altas, cerca de 10 GPa.

Propiedades mecánicas El ZnO es un material relativamente blando con una dureza aproximada de 4.5 en la escala de Mohs. Sus constantes elásticas son menores que aquellas de los semiconductores relevantes de los grupos III-V. La alta capacidad calorífica y conductividad térmica, baja expansión química y alta temperatura de fusión del ZnO son propiedades benéficas para las cerámicas. Entre los semiconductores tetraédricamente unidos, se ha fijado que el ZnO tiene el mayor tensor piezoeléctrico. Esta propiedad lo hace un material de importancia tecnológica para diversas aplicaciones piezoeléctricas, las cuales requieren una amplia unión electromecánica.

Eléctricas el material ZnO tiene un “band gap” relativamente amplio de próximamente $3.3 eV$ a temperatura ambiente. Las ventajas asociadas a un amplio “band gap” incluyen altos voltajes de descomposición, habilidad de mantener amplios campos eléctricos, bajo ruido electrónico, y altas temperaturas y altas potencias de operación. El band gap del ZnO puede ser sintonizado a $3-4 eV$ mediante su aleación con óxido de magnesio u óxido de cadmio.

Casi todo el ZnO tiene el carácter tipo n , incluso sin agregar impurezas intencionalmente. La no estequiometría es típicamente el origen del carácter de tipo n , pero el tema sigue siendo controversial. Una explicación alternativa ha sido propuesta, basada en cálculos teóricos, la involuntaria sustitución de las impurezas de hidrógeno son las responsables. La adición de impurezas de tipo n es controlable cuando se substituye el Zn con elementos del grupo III, tales como Al , Ga , In o al sustituir el oxígeno con elementos del grupo VII como cloro o yodo.

Las limitaciones del dopaje p con límite electrónico y las aplicaciones optoelectrónicas del ZnO , que generalmente requieren uniones con materiales del tipo n y p . Los dopantes conocidos del tipo p incluyen a los elementos del grupo I Li , Na , K ; elementos del grupo V N , P y As ; así como cobre y plata. Sin embargo, muchos de estos elementos son receptores y no producen conducción tipo p significativa a temperatura ambiente.

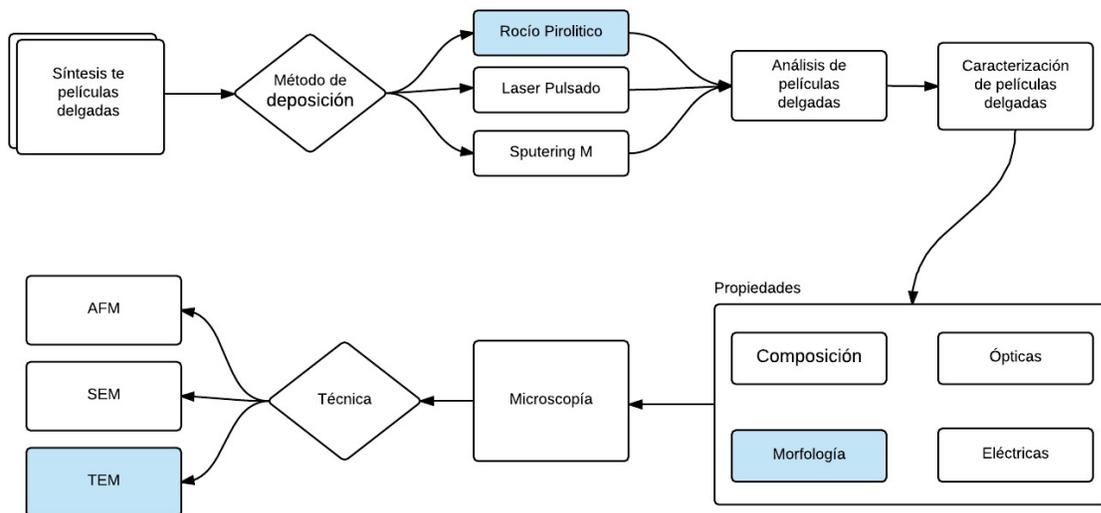


Figura 2.3: Diagrama de flujo de caracterización

1.7 Microscopía

Hoy en día existen muchas variaciones desde el invento del microscopio, que nos permiten ver espacios demasiado pequeños para el ser humano, los microscopios han sido usados para ver, observar, medir, analizar imágenes de seres vivos o estructuras inorgánicas más allá de lo normal.

Hay dos categorías principales en la adquisición de imágenes microscópicas en donde la magnificación es suficiente resolución; microscopios de luz y microscopios de electrones. En el microscopio de luz la magnificación es obtenida por un sistema de lentes y luz, mientras que el microscopio de electrones usa rayos de electrones para producir imágenes. Hay dos tipos de microscopios de electrones *TEM* y *SEM*.

La microscopía óptica es una herramienta utilizada ampliamente en diversas áreas científicas y que cuenta con múltiples aplicaciones en distintos campos como en la biología celular, la inmunología, la medicina tanto humana como animal, la biotecnología y nanotecnología entre otras áreas en que ha incursionado en la actualidad. A medida que crecen los desafíos científicos, este tipo de microscopía se ha acoplado respondiendo a tales necesidades. Por ejemplo, para estudios de membranas celulares se ha optimizado la microscopía de fluorescencia de Reflexión Interna Total (TIRF), que resuelve necesidades específicas.

La microscopía óptica, también denominada fotónica ya que utiliza luz que pasa por un lente para visualizar la muestra, es una herramienta que además de permitir la visualización de muestras ya fijadas, permite la observación de especímenes vivos, haciendo posible la realización de vídeos y por lo tanto, la descripción de procesos en tiempo real. Los avances en la tecnología de los láseres y la digitalización de imágenes entre otros, han hecho que a través de la microscopía óptica logremos evidenciar mecanismos celulares y dinámicas moleculares en tiempo real. Es así como la microscopía óptica, más que ser un instrumento utilizado únicamente para capturar una imagen a manera de foto o sección óptica, conduce a la obtención de datos desde un proceso de indagación más completo.

El microscopio de escaneo de electrones SEM por sus siglas en inglés, funciona con los mismos principios que el microscopio de luz, pero utiliza electrones en lugar de luz. Lo que se puede ver con un microscopio de luz está limitado por la longitud de onda de la luz. Los SEM utilizan electrones como “fuente de luz” y su longitud de onda mucho más baja hace posible obtener una resolución mil veces mejor que con un microscopio óptico. Por ejemplo, puede estudiar pequeños detalles en la celda o dife-

rentes materiales hasta niveles atómicos cercanos. La posibilidad de grandes aumentos ha hecho de la SEM una herramienta tanto en la investigación médica, biológica y de materiales.

Una “fuente de luz” en la parte superior del microscopio emite los electrones que viajan a través del vacío en la columna del microscopio. En lugar de lentes de vidrio que enfocan la luz en el microscopio de luz, el SEM utiliza lentes electromagnéticas para enfocar los electrones en una viga muy delgada. El haz de electrones viaja a través de la muestra que desea estudiar. Dependiendo de la densidad del material presente, algunos de los electrones se dispersan y desaparecen de la viga. En la parte inferior del microscopio los electrones no dispersados golpean una pantalla fluorescente, lo que da lugar a una imagen en la sombra de la muestra con sus diferentes partes mostradas en variada oscuridad de acuerdo a su densidad. La imagen puede ser estudiada directamente por el operador o fotografiada con una cámara.

2 Ciencias de la computación

Áreas importantes de la inteligencia artificial, el aprendizaje profundo ha tenido gran éxito en los últimos años como tecnología detrás de sistemas inteligentes. Las redes neuronales profundas son inspiradas en el modo de como los humanos aprendemos, trabajo de las neurociencias y ciencias de la computación.

2.1 Inteligencia artificial

La inteligencia artificial fue desarrollada a principio de los 50s, es la forma de hacer que las computadoras piensen o actúen de forma inteligente, aplicaciones de la inteligencia artificial visión por computadora, procesamiento de lenguaje natural, reconocimiento de habla, sistemas expertos.

Aprendizaje automático o *Machine learning* y reconocimiento de patrones basados en modelos que aprenden, se realizan predicciones acerca de datos desconocidos, uno de los inconvenientes es que la capacidad del modelo esta relacionada con el poder(calidad) y la cantidad de los datos.

La inteligencia artificial *IA* se a convertido a un tema importante en el mundo, supervisado, cada dato de entrenamiento es asociado a varias características de entrada. Clasificación podemos pensar el problemas de clasificación que en lugar de predecir un variable de valor continuo se predice un etiqueta de clase asociado con el vector de

características.

El aprendizaje profundo actualmente puede proveer soluciones que hasta hace poco se creían resueltas solo por humanos, visión por computadora, procesamiento de lenguaje natural, reconocimiento de voz. Que ha cambiado, grandes cantidades de información disponibles, precios bajos en hardware tanto en discos duros y procesadores, mejores *GPUs*, ardua investigación de nuevos modelos y técnicas.

El aprendizaje profundo es una rama de aprendizaje automático que utiliza datos para saber como realizar ciertas tareas, aprendizaje automático surgido como herramienta para resolver problemas cognitivos, de percepción de los sentidos humanos, el estado del arte visión por computadora y reconocimiento de voz, el nombre de profundo proviene de el uso de redes neuronales profundas compuestas por muchas capas ocultas. Cada capa contiene nodos, que están conectados con la capa anterior como entradas y como salidas en la capa siguiente, son los nodos la estructura de datos donde ocurren los cálculos. Las conexiones entre los nodos se le llaman pesos, parámetros del modelo que son los que se modifican durante el entrenamiento y ajustan para dar el resultado correcto del mapeo de entradas a salidas correctas.

2.2 Aprendizaje Automático

El aprendizaje automático es un subcampo de la informática que evolucionó a partir del estudio del reconocimiento de patrones y la teoría del aprendizaje computacional en Inteligencia Artificial (IA). Arthur Samuel, un pionero en el campo de aprendizaje automático e inteligencia artificial ha definido el aprendizaje de la automático como un *“campo de estudio que da a los ordenadores la capacidad de aprender sin ser explícitamente programado”*.

El aprendizaje automático es un campo de la informática que utilizando métodos estadísticos para crear programas que mejoren rendimiento en el tiempo, o detectar patrones en cantidades masivas de datos que los seres humanos serían poco probable encontrar. El aprendizaje automático explora el estudio y la construcción de algoritmos que pueden aprender y hacer predicciones sobre los datos, tales algoritmos funcionan construyendo un modelo del ejemplo a fin de hacer predicciones o decisiones basadas en datos, en lugar de seguir estrictamente las instrucciones del programa estático.

El aprendizaje supervisado puede ser descrito de manera general como un problema de extraer valor de datos no etiquetados de los cuales existen en cantidades enormes. En cuanto al aprendizaje no supervisado el objetivo es descubrir estructura en los datos

o alguna representación útil, como generarlos, como agruparlos, reducción de dimensionalidad, etcétera.

En resumen, Aprendizaje automático es una colección de algoritmos y técnicas utilizadas para crear sistemas computacionales que aprenden de los datos con el fin de hacer predicciones e inferencias. Los principales tipos de aprendizaje automático son: aprendizaje supervisado, aprendizaje no supervisado, aprendizaje semi-supervisado, aprendizaje por refuerzo.

2.3 Aprendizaje supervisado

los datos de entrenamiento están etiquetados. de igual manera, los problemas de aprendizaje supervisado se dividen en dos, clasificación y regresión. en clasificación la variable a predecir \mathbb{N} es discreta, es decir a una categoría. En el caso de regresión la variable es continua $y \in \mathbb{R}$. Por ejemplo en problemas de clasificación (Reza and Iftekharuddin, 2013) y problemas de regresión (Bremer, 2013).

2.4 Aprendizaje no supervisado

Al contrario de aprendizaje supervisado $X = x, y$, el aprendizaje no supervisado no cuenta con etiquetas, no tiene un valor entero o continuo a predecir, son datos X se centra en encontrar estructura y relaciones entre las diferentes características. Entre las tareas más importantes del aprendizaje no supervisado es la agrupación *clusters* de un conjunto de entradas, además de la agrupación existen otras técnicas; aprendizaje *Hebbiano*, aprendizaje de modelos ocultos, descomposición de valores singulares, análisis de componentes principales, análisis de componentes independientes, etc.

Como parte de la estructura de los datos, es posible aprender la distribución de los datos también llamada “*density distribution*”, una vez teniendo la distribución de los datos se pueden crear muestras que sigan dicha distribución “muestrear”, así mismo generar nuevas estancias siguiendo los datos de entrenamiento.

2.5 Aprendizaje semi-supervisado

Es un mezcla de aprendizaje supervisado y no supervisado, incluye problemas de clasificación cuando no se cuenta con todas las etiquetas, o cuando las etiquetas son de apoyo para problemas de aprendizaje no supervisado. Otro caso, las etiquetas se usan indirectamente en un proceso de aprendizaje supervisado.

2.6 Aprendizaje profundo

Aprendizaje profundo o “*deep learning*” es un subcampo del aprendizaje automático, consiste en una serie sucesiva de capas de representaciones de datos. La palabra profundo en aprendizaje profundo se refiere a el número de capas usadas en el modelo. El aprendizaje profundo ha tenido gran éxito en los últimos años como tecnología detrás de sistemas inteligentes. Las redes neuronales profundas son inspiradas en el modo de como los humanos aprendemos (LeCun et al.).

No obstante, la mayoría de los fundamentos teóricos fueron desarrollados durante los años 1960-1995, la arquitectura canónica de redes neuronales “*feed-forward networks*” fue de los primeros modelos en ser entrenados por algoritmos de entrenamiento llamado *backpropagation*. A partir de este modelo diversas arquitecturas y topologías de redes neuronales han sido desarrolladas; redes neuronales Recurrentes, redes Neuronales convolucionales, máquinas de *boltzmann*, redes codificadoras-decodificadoras.

2.7 Redes convolucionales

Redes convolucionales son redes neuronales que comparten parámetros a través de mapas de características, usualmente los datos son espaciales. Una imagen puede ser representada como un conjunto de capas aplicadas o cubos, que tiene alto ancho y profundidad o número de capas de color. En el caso de imágenes a escala de grises la profundidad es uno. Las capas de convolución se aplican una tras otra. En terminología convolucional las capas ocultas son llamadas *feature maps* o “*kernels*” que son usados para mapear a k mapa de características. Salto o “*stride*”s es la cantidad de píxeles en los que el filtro de convolución (k, k) se mueve sobre la imagen en saltos de s píxeles.

La característica principal de las redes neuronales convolucionales es la operación de convolución, tomada de área de procesamiento de señales, en realidad se trata de una *correlación transpuesta*. Además, el uso de filtros ayuda a problemas con datos espaciales, es decir que sean imágenes, el campo receptivo de los filtros ayudan al modelo a ser robusto a ser invariante a posición, invariante a tamaño, extracción automática de características (filtros) y reducción de parámetros (pesos) a optimizar. En las arquitecturas de redes convolucionales utilizadas para clasificar imágenes podemos identificar dos fases importantes, extracción de características y clasificación, la primera consiste en capas de convolución, cuyo principal fin es descubrir los mejores filtros que clasifiquen mejor al conjunto de entrenamiento mediante operaciones de convolución (figura 2.4), la segunda fase ordena los filtros aprendidos utilizando una operación comúnmen-

te llamada de aplanado, después los mezcla con una red neuronal canónica totalmente conectada y en el proceso final alguna función de transformación probabilidad como la *softmax*, es decir la probabilidad de pertenencia a cada clase.

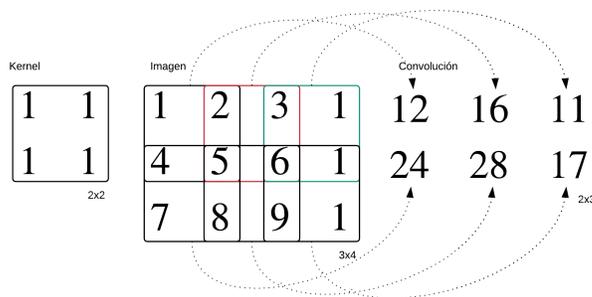


Figura 2.4: Operación de convolución

2.8 Deconvolución

También llamada convolución transpuesta, en muchos casos de visión artificial, la resolución final de las imágenes requieren ser mayor que las entradas, esta capa es usada en modelos como *DCCGANs*, super resolución de imágenes, estimación de profundidad. las *CNN* realizan operaciones que modifican la imagen *downsampling* producen que la salida en cada capa sea menor dimensión en alto y ancho que las anteriores, es decir menor dimensión. En cuanto a la convolución transpuesta realizan un *upsampling*, lo opuesto a lo anterior. La deconvolución es un termino mal empleado porque también realiza operaciones de deconvolución.

2.9 Modelos generativos

En aprendizaje automático, un modelo generativo es aquel que tiene como objetivo crear nuevas instancias con modelos de distribución $p(x)$, describe como los datos son generado probabilísticamente, las imágenes son por ejemplo los datos más usados por modelos generativos. En (Pfau, 2017) dice que los usos principales de redes generativas adversarias son experimentos científicos y arte.

2.10 Auto-codificadores

(Baldi, 2012), *autoencoders* es el tipo más simple de una red neuronal para problemas no supervisado, el número de nodos en la capa oculta de la red es siempre menor a la capa de entrada, el objetivo es comprimir los datos de manera eficiente. la capa de salida reconstruye la capa de entrada $fit(X, X)$. 1 capa oculta. reducción de dimensionalidad, compresión de datos. permite construir modelos generativos de donde muestrear. Otra aplicación es la de pre-entrenamiento no supervisado para entrenamiento supervisado.

2.11 Auto-codificadores variacionales

textitvariational autoencoders (Kingma and Welling, 2013), se definen como modelos gráficos probabilísticos que tienen como objetivo principal en modelado de variables latentes es una red neuronal que aprende a reproducir sus entradas, mapea los datos de entrenamiento a un espacio oculto. y luego generar muestras de la distribución de los datos. Son un tipo de modelos generativos que muestran desde espacio oculto, se pueden resumir como una combinación de redes neuronales y paradigma bayesiano.

La palabra variacional se refiere a inferencia variacional o bayes variacional, cae el el termino de *bayesian machine learning* aprende distribuciones desde z “*latent variable*”-representación de variable oculta z capa oculta. Los problemas en bayes se tratan de aprender distribuciones no puntos de estimación. Encontrar $q(z)$ que nos dice que PFD z será, se puede transformar la salida de encoder en media y varianza.

2.12 Redes generativas adversarias

(Goodfellow et al., 2014) Son un tipo de aprendizaje profundo que aprende el modelo de distribución de los datos de entrada (dataset) $p_{model}(x)$, las *GANs* (del inglés “*Generative Adversarial Networks*”) tienen una arquitectura de dos redes neuronales con funciones de; generador G que crea imágenes del dataset siguiendo una distribución de probabilidad y el discriminador que aprende a distinguir el origen de la imagen, si proviene de datos reales (la imagen) o proviene del generador. Se convierte en una prueba de imágenes generadas (falsa)contra reales y termina cuando el discriminador no es capaz de determinar si una instancia x proviene de los datos reales o fue generado artificialmente con el generador G .

la función objetivo $V(G, G)$ resulta en la distribución de probabilidad del generador sea exactamente igual a la distribución de los datos reales. esto significa que las muestras falsas no sean distinguidas para el discriminador. Las GANs se centran en aprender la distribución de los datos,

Las *GANs* producen mejores muestras que *VAEs*, pero no hay un número que lo diga, la apreciación es cualitativa. la idea general es de un juego de redes neuronales que una G generador trata de burlar el discriminador D , puede hacer esto al generar muestrear cada vez mejores, ejemplo: falsificador de billetes. Las redes gans no tienen componentes especiales.

Función costo trata de minimizar y la otra maximizar simultáneamente, tratan de optimizar cosas opuestas. El discriminador recibe dos tipos de imágenes real y falsa. es decir una clasificación binaria, *binary cross entropy*, *BCE*, $J = -[t \log y + (1 - t) \log (1 - y)]$ t representa el 1 y y representa la probabilidad de salida del discriminador.

Los pasos para generar una muestra desde $p(x)$.

1. $z \sim p(z)$: z sigue una distribución normal.
2. concatenar z y x como entradas para G
3. $\hat{x} = G(zx)$: muestrear con parámetros de G es $G(z; \theta_G)$.

Juego minmax El entrenamiento puede ser visto como un juego minimax de la teoría de juegos, las dos redes compiten una contra la otra donde $G(x)$ su objetivo es maximizar y $D(x)$ a minimizar (es un clasificador), el objetivo general es alcanzar el juego óptimo, un punto de equilibrio entre las dos redes, la salida de $G(x)$ son imágenes generadas, y $D(x)$ un rango de probabilidad $[0 - 1]$ que corresponda a clasificar la salida de $G(x)$ como 0 o 1 donde 0 se considera muestra falsa y 1 real que proviene del conjunto de entrenamiento. Al alcanzar el punto perfecto de entrenamiento, el punto de equilibrio, el $D(x)$ no es capaz de diferenciar de donde proviene la salida $G(x)$ entonces el rendimiento de $D(x)$ es igual a 0.5 o 50 de instancias correctamente clasificadas, debido a que las muestras generadas de $G(x)$ y el dataset son muy parecidas que el discriminador no puede diferenciar x . Algunas veces el modelo no alcanza el equilibrio por problemas en el entrenamiento, es importante el entrenamiento de G y D sea al mismo tiempo, entre otros parámetros.

DCGAN

Deep convolutional generative adversarial networks ([Radford et al., 2015](#)), a partir de este trabajo las GANs son utilizadas con técnicas para generar imágenes, utilizan

redes conolucionales para clasificar y generar, hace contribuciones significanticas al rendimiento de GANs y arquitectura. Remplazan las capas de pooling usando saltos convoluciones "*strided convolution*", que sea habían usado para reducir la complejidad computacional del modelo (dimensiones), incrementar el campo receptivo, lo que resulta en a ser invariante a escalamiento. Quitar las capas totalmente conectadas y conectar la salida directamente con la anterior para convolucional.

La normalización por lotes rescala la entrada de cada capa para que tenga media de cero y uno de varianza. Ayuda a el modelo en entrenamiento evite el colapso en la media "*mode collapse*". Cuando no se aplica en modelo también tiende a un entrenamiento inestable. el uso de la función de activación *ReLU* en todas las capas después de la convolución del generador, excepto en la última utiliza *tanh* , así mismo en las capas de discriminador con función de activación *leaky ReLU*.

El discriminador es esencialmente una imagen en espejo del discriminador. he 100-D noise input is fully connected the high level convolutional features. This layer then uses fractional-striding to double the size of the filters, but creates half the number. This process of doubling in size, halving the number is repeated until 128 filters of size 32x32 are created. This is then upscaled to a 64x64 image with 3 layers, representing the three color channels.

Los hiperparámetros como *leaky ReLU* son producto de experimentos e investigación. Trabaja mejor en las mismas capas de convolución como convoluciones trasnpuestas y estabiliza el entrenamiento con las siguientes modificaciones: remplace las campos de pooling con convoluciones en el discriminador *fractional stride convolutions*, usa normalización por lotes in el generado y también en el discriminador, usa activación ReLU en el generador para todas las capas excepto en la de salida la cual usa $\tanh x$, usa la función de activación leaky ReLU en el discriminador en todas las capas. <https://github.com/jacobgil/keras-dcgan>.

En la academia es normal bautizar las GANs por su área de investigación, mejoras, aplicación etc. Pero esencialmente son el mismo concepto anterior, en Wasserstein GAN (wGAN) (Arjovsky et al., 2017) más detalles en sección 2.13, lo podemos encontrar en trabajos como (Zhang et al., 2016).

2.13 Función costo

La función costo a minimizar está en términos de distribuciones de probabilidad, las formas de comparar dos distribuciones X y $X G(z)$ son **Kullback–Leibler y Jensen–Shannon Divergence**. Divergencia Kullback-Leibler o divergencia KL es utiliza-

da para medir la diferencia entre dos distribuciones de probabilidad sobre una variable x , el concepto fue originado en teoría de la probabilidad y teoría de la información. La divergencia KL esta relacionada con términos como entropía relativa, divergencia de la información, es una medida no simétrica de la diferencia entra dos distribuciones de probabilidad ejemplo, $p(x)$ y $q(x)$. en la notacion de la divergencia KL de $q(x)$ sobre $p(x)$ se denota $D_{KL}(p(x), q(x))$ es una medida de pérdida de formación cuanto $q(x)$ es usada para aproximar a $p(x)$, comúnmente $p(x)$ representa la distribución real de los datos, las observaciones directas, la medida en q representa una teoría, un modelo, una descripción o en nuestro caso un aproximación a $p(x)$. La distancia D_{kl} es cero cuando las probabilidades p y q son iguales.

La Divergencia Jensen-Shannon (Divergencia JS), es un método de similaridad entre dos distribuciones de probabilidad, está basada en la divergencia KL con la diferencia de que es simétrica $D_{JS}(p(x), q(x)) = D_{KL}(q(x), p(x))$ y tiene un valor finito y el valor de distancia Jensen-Shannon está definida como la raíz cuadrada de la divergencia JS.

La métrica Wasserstein es una función de distancia definida en términos de distribuciones de probabilidad en un espacio M , Wasserstein mide el costo de trasportar la masa de $q(x)$ a $p(x)$, *Wasserstein*, *Heart-mover* se utiliza en *wGAN*.

La función costo 2.1 es una unión de el costo de ambas redes G y D , la parte derecha de el signo $+$ corresponde a D y la parte izquierda corresponde a D .

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x, y)] + \mathbb{E}_{z \sim p_z(x)} [\log (1 - D(G(z), y))] \quad (2.1)$$

En la ecuación 2.1, El discriminador entrena para maximizar la probabilidad de asignar la etiqueta correcta para las muestras reales y las generadas, mientras tanto el generador está entrenando para minimizar $(\log 1 - D(G(z)))$ minimizar la probabilidad de que el discriminador acierte a la etiqueta de $G(z)$, en otras palabras confundir a D es bueno para G .

2.14 Comparación de modelos generativos

Es importante diferenciar los diferentes modelos generativos y otros modelos gráficos con sus ventajas, los auto-codificadores, codifican las entradas en un espacio menor, después realiza una decodificación o reconstrucción de los datos, son de tres capas entrada, salida, oculta son útiles encontrando una representación compacta de la distribución

de los datos. Auto-codificados variacionales son una extensión de los auto-codificados con enfoque bayesiano, representaciones ocultas del vector latente z , son parámetros de distribuciones normales, apartir de el tamaño del vector latente se podrán generar imágenes menos borrosas. Los VAEs sufren de generaciones de imágenes borrosas al no ser capas de aprender detalles de la distribución de datos, lo anterior se debe el proceso de aproximación de parámetros de la distribución, que usualmente es una distribución compleja.

En caso de las gans el generador puede aprender la distribución de datos indirectamente con “pistas” del discriminador, estas pistas son el gradiente de la función objetivo que minimizan la similitud de la distribuciones de los datos y distribución de las imágenes sintéticas, con métricas como divergencia KL y Divergencias JS o la distancia wasserstein-1.

En recientes desarrollos han mezclado estas ideas en modelos como (Makhzani et al., 2015) autoencoders adversarios, también con enfoque bayesiano (Kingma and Welling, 2013) , modelos con funciones de energía (Zhai et al., 2016).

2.15 Redes generativas adversarias condicionales

Las redes generativas adversarias condicionales pueden verse como una extensión del modelo original si el discriminador y el generador son condicionados a información extra. En **GAN** No hay variables, atributos que explícitamente controlen el generador así que el generador no está condicionado es libre de generar cualquier muestra de $G P(z)$. a continuación algunos ejemplos.

CGAN (Mirza and Osindero, 2014) Un atributo es representado usando una etiqueta observable c , con una variable binaria indicando la presencia del atributo c , el entrenamiento está restringido al valor de la etiqueta en las entradas de el generador y discriminador. En el generador el vector de ruido $p_z(z)$ y c son concatenados en una representación z, c , $X G(X|z, y)$ ahora el discriminador intento modelar $d D(d|X, y)$. Función objetivo en la ecuación 2.1.

CFGAN (Kaneko, 2017) tiene una arquitectura de filtro f_y está incorporada en la entrada del generador, el filtro asocia el atributo y con la variable latente multidimensional z_a . con esto el generados puede ser controlado multidimensionalmente.

InfoGan (Chen et al., 2017) la principal propuesta de la arquitectura GANs es poner junto a las entradas de generador z una entradas c salientes. estas variables latentes salientes pueden ser muestreadas desde cualquier distribución uniforme, gaussiana etc.

El discriminador intenta recuperar las variables latentes con solo mirar la imagen generada, la cual es usada para formar un término “regulación de información” dentro de la función objetivo de la GAN. en lugar de solo poner en la salida de la predicción si es real o falsa. El discriminador también tiene una salida \hat{c} para la reconstruir c .

A continuación la figura 2.5 muestra la relación jerárquica de donde se sitúan las redes condicionales adversarias respecto a otros enfoques de aprendizaje automático.

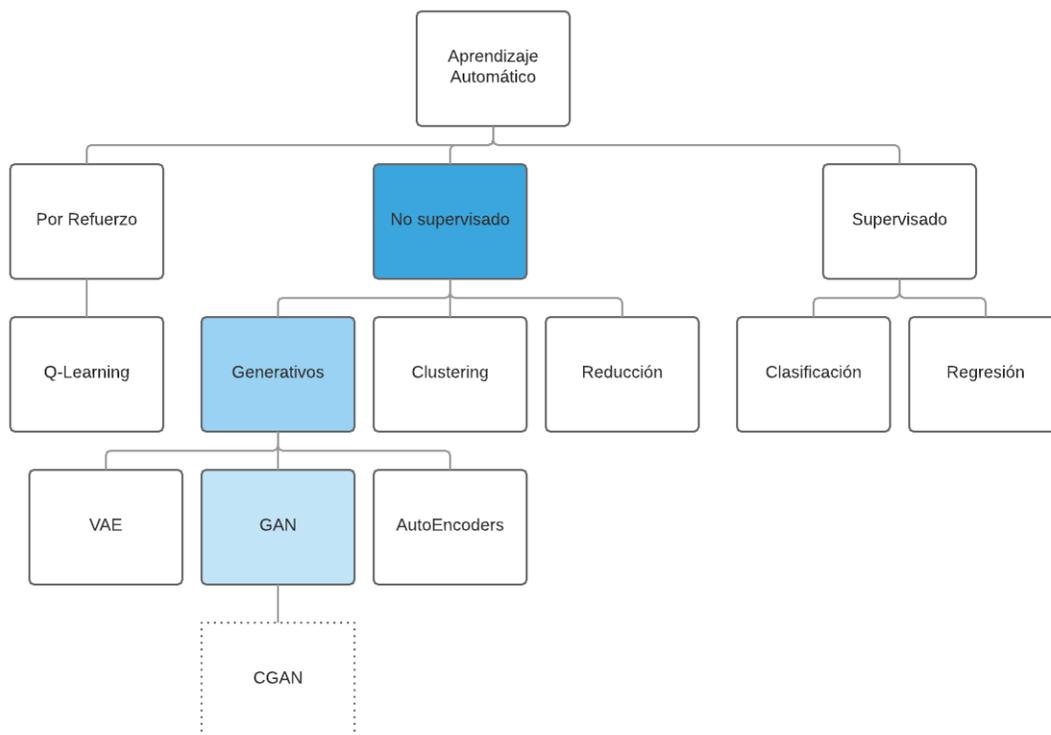


Figura 2.5: Diagrama de sub-área en GAN

2.16 Arquitectura de las redes generativas

Los desarrollos iniciales de redes generativas adversarias no describen como debe de ser el agente generador y discriminador, son implementadas por redes neuronales artificiales totalmente conectadas. Los avances recientes involucran avances en aprendizaje profundo, lo que concluyó a que el generador está constituido por capas de convolución transpuesta y discriminador está conformada por capas de convolución.

Las arquitecturas iniciales. El presente trabajo se desarrollo bajo la arquitectura *deep convolutional generative adversarial networks* o *DCGAN* (Radford et al., 2015), a continuación se describen los componentes más destacados que son utilizados en el presente trabajo.

Discriminador el trabajo principal de la red discriminador D es solo clasificar la entrada, es un clasificador binario(dos clases), predice 0 si la entrada proviene de los datos sintéticos del generador(salida). la función está definida como 3.7 conformado con la siguiente estructura.

- **Dimensión de entrada:** tamaño de imágenes 64×64
Número de capas: cuatro capas de convolución
- **Dimensión de salida:** un escalar de $[0, 1]$
- **Término de *bias*:** sin vector parámetro de bias θ_0 en ninguna capa
- **Saltos *stride*:** 0
- **Margen *padding*:** 0
- **Dimensión del *kernel*** dimension de la ventana de convolucion 4×4
- **Función de activación:** sigmoide
- **Inicialización de parámetros** con distribución normal truncada

Generador la red generativa G genera imágenes a partir es un vector oculto de variables z donde realiza una serie de convoluciones transpuestas hasta alcanzar la dimensiones del tensor deseados. la operación transpuesta realmente lo que hace es escalar en resolución (dimensión) la entrada de los pixeles para generar imágenes. El proceso de aprendizaje ocurre cuando el discriminador clasifica como real la imagen, es decir “burla” a el clasificador.

- **Dimensión de entrada:** vector de variables latente z y c
Número de capas: cuatro capas de convolución traspuesta
- **Dimensión de salida:** 64×64
- **Término de *bias*:** sin vector parámetro de bias θ_0
- **Saltos *stride*:** 0
- **Margen *padding*:** 0

- **Dimensión del *kernel***: dimension de la ventana de convolucion 4×4
- **Función de activación**: hiper-tangencial
- **Inicialización de parámetros**: con distribución normal truncada.

3 Métricas de evaluación de modelos

La evaluación cuantitativa de los modelos generativos adversarios. inception score, MODE score, AM score.

Inception score (Bethge, 2016) es una métrica propuesta para evaluar el rendimiento del generador, el *inception score* demostró tener una buena correlación con evaluaciones humanas, donde un *inception model*¹ Pre-entrenado está disponible públicamente (Szegedy et al., 2014) se evalúa a cada imagen generada x y obteniendo la clase de distribución de la probabilidad en las clases C . Además en trabajos recientes otras alternativas se han desarrollado como son *MODE score* (Pfau, 2017) y *AM score* (Zhou et al., 2017), la primera mide la relación de moda a cada muestra generada y la segunda evalúa la probabilidad de las muestras generadas sean reales.

3.1 Trabajos relacionados

La conjunción de estas dos ramas de la ciencia y tecnología es relativamente nueva, los esfuerzos se han exponenciado en la última década y la clasificación de imágenes es una tarea del aprendizaje automático más utilizada, describe una función f que mapea las entradas (características) a una etiqueta, entre los trabajos relacionados más destacados ordenados en forma ascendente en el tiempo:

(Varde et al., 2006) utilizan minería de datos y procesamiento de imágenes para crear datasets científicos bien organizados y categorizados por forma, textura y poder medir similitudes.

¹ Arquitectura de Google inc. ganadora en el concurso ILSVRC edición 2014 originalmente llamada GoogleLenet. Introduce un concepto de modulo de *inception*, que transforma las capas de convolución en un solo bloque de mapa de características por medio de operaciones de convolución. Tiene mejoras en rendimiento de memoria y velocidad.

(Suh et al., 2010) utiliza técnicas de minería de datos para correlacionar base de datos multidimensionales de matrices de reacciones fotovoltaicas en óxidos transparentes conductivos utilizando óxido de ZnO como semiconductor.

(Ebrahimi et al., 2012) enfoque bayesiano para la predicción de nano estructuras en sustratos de $CdSe$, desarrollaron un modelo de optimización basado con enfoque bayesiano para los parámetros experimentales necesarios para sintetizar las morfologías deseadas. Este artículo demuestra que métodos bayesianos pueden ser útiles en analizar datos experimentales y asistir el proceso de identificación de óptimas condiciones para hacer crecer la nanoestructura

(Shehadeh et al., 2012) muestra un trabajo de clasificación de morfología de ZnO en nanoestructuras; *nanobelts*, *nanowires*, *nanosaws* mediante diferentes técnicas de minería de datos, utiliza parámetros de deposición de nanoestructuras para predecir la morfología.

(Gauthier, 2014) aplican una extensión de $GANs$, llamada $cGANs$ condicionales para la generación de rostros, En $GANs$ el generador intenta engañar al discriminador haciéndole creer que las imágenes que genera son reales, en tanto el discriminador clasifica las imágenes como falsas o verdaderas. la diferencia está en el generador es provisto de información adicional que resulta que la red generador genere rostros con atributos en específico.

Las redes neuronales generativas también pueden ser usadas en simuladores como generador de mapas cosmológicos. (Mustafa et al., 2017), su estudio abarca estudios estadísticos sobre su creación.

(Osokin et al., 2017) aplican redes generativas adversarias para la síntesis de imágenes de microscopía fluorescente. Generan imágenes por canal de color separado. Interpolando variables a través del espacio oculto, pueden simular cambios en la lugar de proteínas en la célula durante su ciclo de vida. Los resultados cuantitativos son descritos con una técnica llamada *“inception score”*.

Con el uso de las redes generativas adversarias (Nguyen et al., 2017) entrenan modelos para la generación para representación de sombras y las relaciones entre los objetos en la escena mediante un parámetro de sensibilidad. Un detector de sombras que genera un mapeo de sombras de una imagen de entrada.

Análisis y diseño de caracterización de nano estructuras, (Malkiel et al., 2017) realiza predicciones de la caracterización de propiedades físicas como respuesta del espectro óptico en nano-estructuras utilizando redes neuronales profundas.

Uso de las GANs en otras áreas En la industria farmacéutica: La arquitectura GAN-RL (redes generativas y aprendizaje por refuerzo) (Kadurin et al., 2017) propuesta que demostró la capacidad de generar un porcentaje sustancial de estructuras moleculares válidas y únicas. Este estudio utilizó representaciones de cadena de estructura molecular e internamente estamos utilizando generadores integrados múltiples con refuerzo el aprendizaje y la representación patentada de la estructura molecular, que nos permite sintetizar las moléculas exactas y vincular la química y la biología.

Capítulo 3

Metodología

En este capítulo, se presenta el desarrollo de la tesis, desarrollo del modelo computacional. El presente trabajo tiene como meta simular la experimentación científica en nanoestructuras a través de un modelo generativo introduciendo como parámetros condicionales al modelo, las condiciones de deposición de películas delgadas de óxido de zinc.

1 Descripción de la metodología

EL proceso se lleva a cabo en cuatro fases; Imágenes, pre-procesamiento, entrenamiento del modelo y descripción de resultados, con tareas en específico en cada fase breve descripción de la metodología en la Figura (1.2).

Antes que nada, se obtuvieron las **imágenes** (descritas en **ref**), el pre-procesamiento en tres tareas principales que son; una *selección de imágenes*: de acuerdo a los objetivos de la investigación, *Mejoramiento de la imagen*: Por su naturaleza de digitalización las imágenes pueden tener ruido, *Aumentación de datos*: Se centra en obtener más imágenes a partir de las originales pre-procesadas.

En **entrenamiento del modelo**: *Selección de arquitectura* de acuerdo al estado del arte y experimentos, [quitar selección de modelos], optimización de Hiper-parámetros, de acuerdo la experimentación de entrenamiento seleccionar los mejores parámetros de entrenamiento que dan mejores resultados. Resultados Mostrar imágenes como *resultado* del modelo generativo, así *validar* la generación con métricas específicas.

De manera general el presenta trabajo sigue la metodología representada en la

figura (1.2), a continuación y durante todo el capítulo 3 se describen todos los pasos de manera detallada.

1.1 Adquisición de información

El área de aprendizaje automático es una forma de resolución de problemas, que, a través de las estadísticas puede resolver problemas de clasificación y regresión. Por lo anterior, se dice que las soluciones son enfocadas a los datos, lo que significa que es lo más importante de un algoritmo, según Google *"No poseemos los mejores algoritmos, tenemos grandes cantidades de datos"*.

Una de las tareas más importantes para un científico de datos es la llamada adquisición de información, que consiste principalmente en extraer la información, seleccionar información, pre procesar, análisis exploratorio entre otras. La siguiente sección describe la creación del conjunto de datos utilizado para realizar la clasificación de nano estructuras en imágenes.

2 Descripción de las imágenes

Las imágenes son producto de un arduo trabajo de caracterización de películas delgadas de diferentes instituciones educativas (ver anexo A). El proceso de adquisición de imágenes con películas delgadas de ZnO sobre sustratos de vidrio y depositadas por la técnica de rocío pirolítico ultrasónico, fue con equipo microscopio de barrido de electrones (SEM) y almacenadas en formato digital **.tiff**.

Las imágenes muestran nano estructuras tomadas desde diferentes magnificencias los conjuntos de imágenes suman una cantidad de las propiedades de las imágenes como la resolución difieren en un rango de 1800 pixeles a 1200 pixeles, cuentan con información dentro de la imagen que describe los parámetros de digitalización (figura tal).

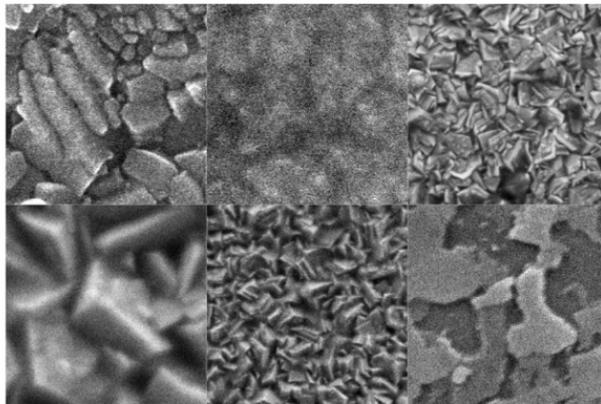


Figura 3.1: Muestra de imágenes (pre-procesadas) del dataset

3 Aumentación de datos y pre-procesamiento

3.1 Selección de imágenes

Debido a que el proceso para digitalizar este tipo de muestras es muy tardado, la cantidad de imagen se considera grande para ser resuelto con modelos aprendizaje profundo que generalice bien. Sin embargo, se considera como una muestra de una población que nos permite hacer inferencias sobre la precisión del modelo.

Un problema fue que la conjunción de áreas de nanotecnología e inteligencia artificial no es un campo muy estudiado, de tal modo que no existen actualmente bancos de datos que nos puedan servir para nuestro objetivo.

El problema es resuelto fácilmente al recortar la imagen de modo a quitar el texto, lo anterior se puede hacer porque las imágenes son relativamente grandes en resolución y no se pierde información relevante. Según (Tolstikhin et al., 2017) la forma más óptima de resolución de las imágenes es cuadrada, lo que facilita la especificación de filtros "kernels", "strides", "padding", etcetera. El ajuste de resolución cuadrada para todas las imágenes que se propone es de 700 alto por 700 de ancho. Lo anterior con el fin de preservar el aspecto del ratio 1:1 y no modificar la imagen de manera drástica.

Cabe destacar que esta primera etapa de procesamiento solo modifica la imagen quitando el ruido y normalizando la resolución. Para finalizar, la resolución final propuesta

es de 200 por 200 píxeles debido a la complejidad algorítmica que implica trabajar con imágenes, es decir, un modelo para imágenes de 28 por 28 llega a ocupar más de un millón de parámetros, por consiguiente, todos los trabajos relacionados con aprendizaje profundo requieren re-escalamiento de imágenes (Simonyan and Zisserman, 2014).

De manera general, el procedimiento llevado a cabo es, seleccionar imágenes de interés dentro del conjunto, seguido de pre- procesamiento y escalado de imágenes, y el último paso es opcional pero muy recomendable [], trata de mejorar el rendimiento gradiente rescalando el espacio de solución a dominio de $[0 - 1]$, se logra modificando el formato de niveles de gris en cada capa de la imagen, que va de 0 hasta 255 a 0 hasta 1. Como resultado de este proceso las imágenes normalizadas listas para ser usadas como entrada de la red neuronal convolucional.

3.2 Pre-procesado de imágenes

El procesamiento de imágenes en una método que realiza operaciones sobre una imagen con el fin de mejorar la para cierto proceso o extraer información útil. Filtrado es una técnica para modificar o mejorar la imagen en un dominio espacial, las operaciones se realizan tomando en cuenta píxeles vecinos al píxel que se analiza la principal fuente de energía para las imágenes es el espectro de energía electromagnética, otros, acústica, ultrasónica, electrónica. Imágenes basadas en radiación espectro electromagnéticas ondas electromagnéticas pueden ser pensadas como un flujo de partículas sin masa que viajan en forma de onda moviéndose a la velocidad de la luz. cada partícula tiene una cantidad de energía llamada fotón. Mejoramiento de percepción humana y representación más conveniente para la computadora.

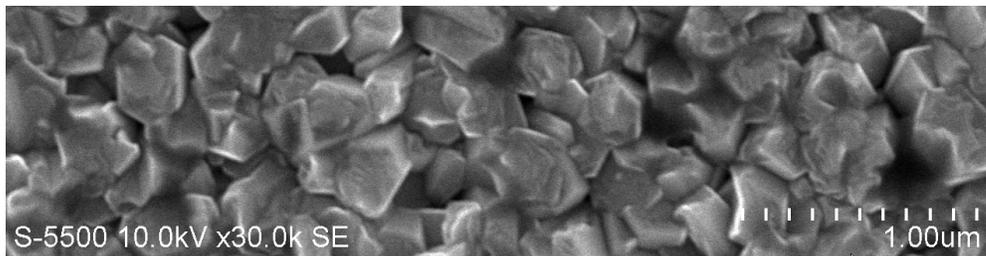


Figura 3.2: Imagen con leyendas

Niveles de procesamiento Los niveles están divididos como escalones que se dirigen al entendimiento semántico de la imagen, niveles de procesamiento

Nivel 1: Operación de primitivas como pre procesamiento de imágenes para reducir el ruido, ajustar contraste, etc. En pocas palabras es el mejoramiento de la imagen para un fin.

Nivel 2: Involucra métodos para extraer atributos de las imágenes y los transforma de una forma mejor manejable para la computadora. Por ejemplo, segmentación, Reconocimiento, extracción de características.

Nivel 3: Se refiere a hacer sentido, emula funciones cognitivas de la visión.

Las imágenes a menudo cuentan con ruido por diversos factores, ya sea por método de digitalización, tipo de imagen, luz, dispositivo utilizados y en este caso etiquetas, texto dentro de la imagen que no interesa (figura 3.2).

Mejoramiento de la imagen

La aplicación de filtro gaussiano, en esta fase de pre procesamiento, la aplicación de filtros para mejorar la calidad de la imagen, ruido de presenta en pixeles dispersos sin relación de contraste en lo pixeles vecinos. Algunas imágenes están con mucho ruido, al aplicar un acercamiento algunas imágenes no tienen una estructura definida, figura 3.3.

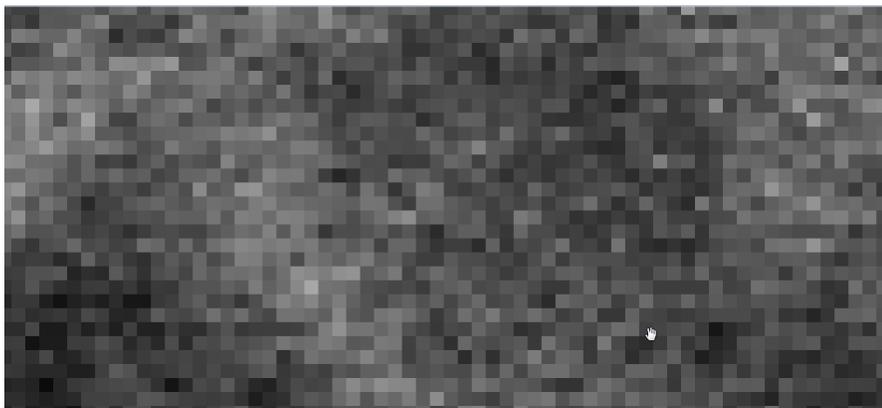


Figura 3.3: Acercamiento a las imágenes

El filtro gaussiano suaviza la imagen quitando esos pixeles mediante un kernel (tamaño de ventana) y una desviación estandar σ , entre más grande el filtro más pixeles entran en la operación y entre más σ el efecto de suavizado se intensifica. Aplicación de diferentes filtros se muestra en la siguiente imagen.

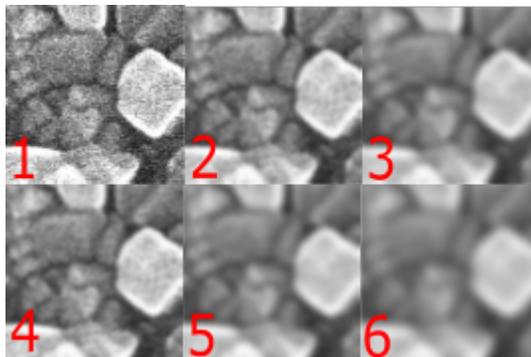


Figura 3.4: Resultados con diferentes filtros

1. Imagen original
2. Con filtro de tamaño 5 desviación estándar 1
3. Con filtro de tamaño 7 desviación estándar 2
4. Con filtro de tamaño 9 desviación estándar 1
5. Con filtro de tamaño 9 desviación estándar 2
6. Con filtro de tamaño 9 desviación estándar 3

Los resultados se muestran en figura 3.4, muestran que un kernel $k = 7$ y $\sigma = 2$ son los que dan mejoras prestaciones entre ruido y efecto borroso. el objetivo principal es quitar el contraste entre pixeles con niveles de gris similar, para obtener como resultado imágenes más claras en los bordes, border mejor definidos que son útiles en los mapas de características de las capas convolucionales.

4 Aumentación de datos

Esta técnica consiste en obtener más datos que con los que se cuenta, existen varias técnicas para diferentes tipos de datos, en nuestro caso los datos son imágenes (Wong et al., 2016). El motivo principal de estas técnicas es que el algoritmo de aprendizaje no “vea” la misma instancia en todas las épocas, si no, modificando la instancia original un poco de manera de que no se pierdan sus características. Lo que hace que el algoritmo generalice mejor sobre instancias no vistas (conjunto de prueba), reduciendo así el sobre ajuste.

Tipos de aumentación de datos en imágenes utilizados.

1. **Multi-Cropping:** Recorta imágenes en coordenadas distintas para obtener nuevas imágenes más chicas pero representan el mismo objeto.
2. **Mirroring:** Modifica la imagen efecto de espejo.
3. **Transformación geométrica:** Modifica la imagen con una ecuación lineal.
4. **Filtros:** Utiliza filtros para agregar "ruido" a la imagen.

Tomando en cuenta la naturaleza de los datos, resulta factible realizar dicha técnica, debido a que las nano-estructuras se repiten a lo largo de toda la imagen en diferentes posiciones y formas. Es posible realizar recortes múltiples con una ventana deslizante sin perder información y de esta manera aumentar significativamente el tamaño del conjunto de entrenamiento.

A continuación, se describen los pasos y parámetros para llevar a cabo el "multi-cropping".

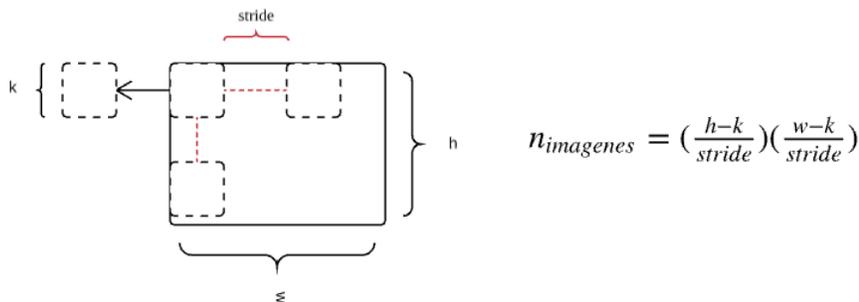


Figura 3.5: Cálculo de filtros

Las imágenes del dataset original tienen diferentes dimensiones y se pueden obtener diferentes ventanas (imágenes recortadas). Los parámetros que influyen en este proceso son: *stride* y *k*, donde *k* es el tamaño de un lado de la ventana, que siempre va a ser cuadrada por simpleza y control [], y *stride*, que es el tamaño de movimiento en píxeles que la ventana realiza paso a paso. Entre mayor sea el *stride* menos imágenes obtenemos, a su vez, entre mayor sea el valor *k* menor es la cantidad de imágenes adquiridas.

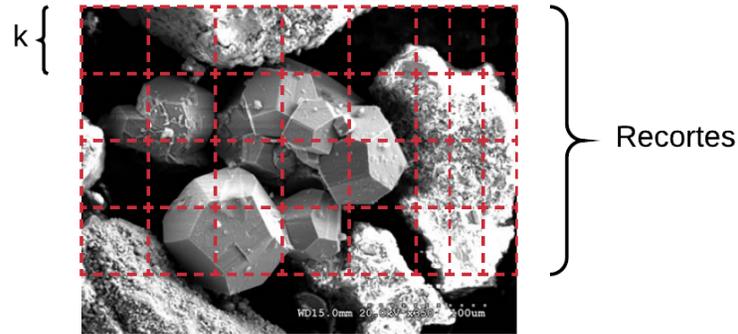


Figura 3.6: Cálculo de filtros

Lo anterior para cada una de las imágenes del dataset, lo que conlleva a un aumento considerable del dataset, el total de imágenes está dada por la siguiente ecuación:

$$numero = \frac{alto - k - 2p}{s} + 1 \quad (3.1)$$

Las imágenes de entrenamiento cuenta con un total de 348, al aplicar esta técnica con $k=200$ y $stride=150$, tenemos como resultado que la cantidad de imágenes ha incrementado 3450.

Estandarización de imágenes las entradas para los algoritmos de aprendizaje automático deben de ser homogéneos en todas las instancias, en este caso la mejor representación de las imágenes no es en valores de [0-255] niveles de gris, es un escalado a [0-1] con una transformación de σ 0.4 para cada pixel en todas las imágenes.

5 Modelado

5.1 Modelo

Función costo Esencialmente la función costo de las *GANs* cuantifica la similitud de distribución entre la distribución del generador P_g y la distribución real de las

muestras P_r , el mejor generador G^* es aquel que replica la distribución de los datos a un costo mínimo de $L(G^*, D^*) = -2 \log 2$

5.2 Arquitectura

De manera gráfica la arquitectura se muestra en figura 3.7, el generador está compuesto por las entradas son el vector aleatorio z y las variables condicionales c , después 4 bloques de deconvolución y uno de aplanado (*flattening*), cada bloque tiene proceso de convolución transpuesta seguido de una normalización y la función de activación ReLU, el objetivo de estos bloques es modificar poco a poco las dimensiones del tensor de entrada zc de dimensión $100z + 4c$ a un tensor de dimension $64(alto) \times 64(ancho) \times 3(canales)$. El discriminador esta conformada por una arquitectura de clasificación convolucional para problemas binarios, con 4 bloques de convolucion que descomponen las representacion de los datos en mapas de características, conforme pasan los datos, la dimension del tensor de entrada $64 \times 64 \times 3$ se va reduciendo a dimension 1, el cual se utiliza directamente para la clasificación 0 imagen falsa, 1 imagen verdadera.

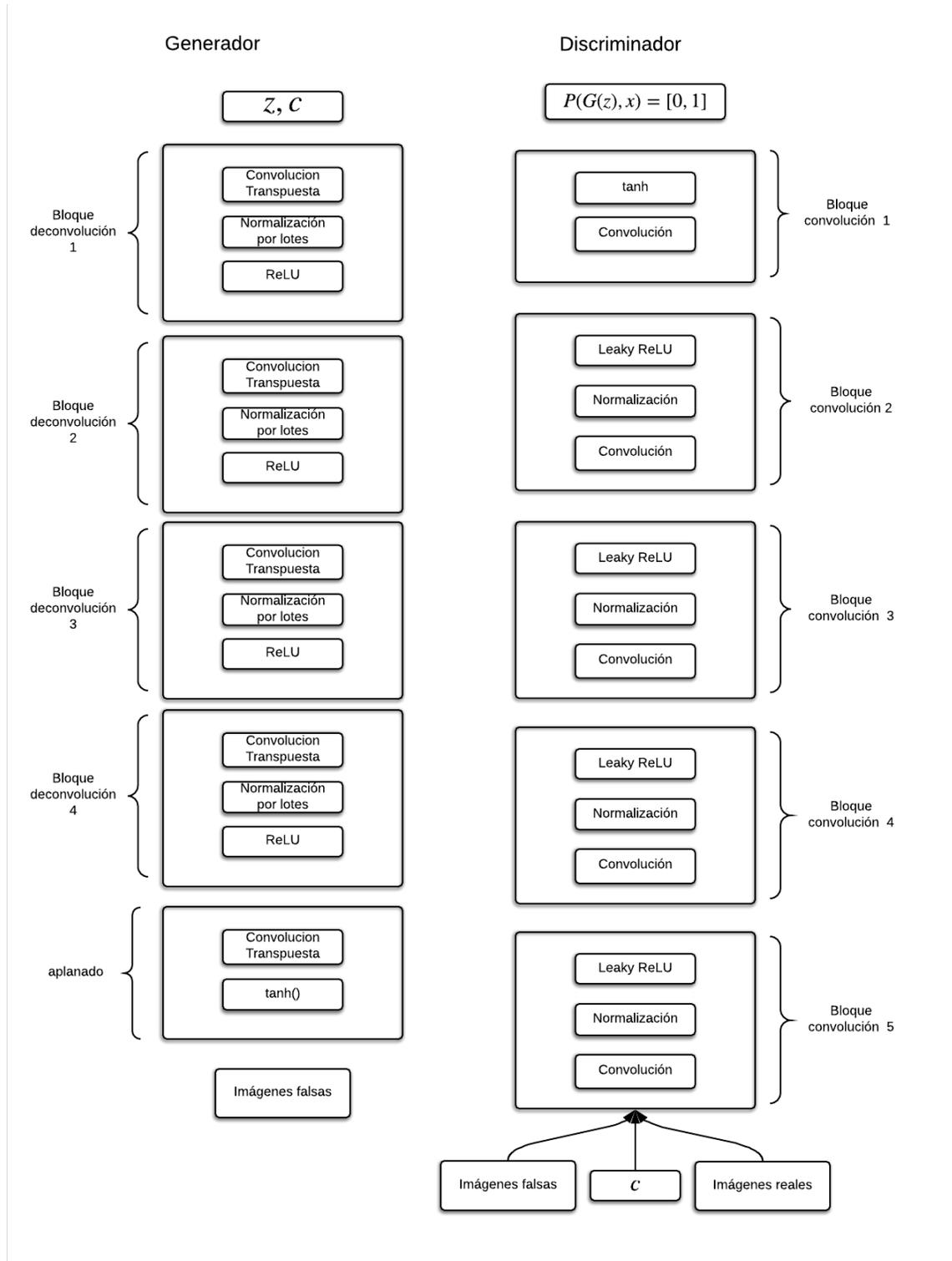


Figura 3.7: Diagrama en bloques de las redes generador G y discriminador D

5.3 Entrenando GANs

Los hiper-parámetros de ambos modelos, generador y discriminador fueron obtenidos de artículos científicos que demuestran técnicas y trucos obtener un entrenamiento estable (Salimans et al., 2016), Wasserstein GAN *wGAN*, que propone algunas mejoras para hacer el proceso de entrenamiento más estable, aunque en algunas circunstancias aún genera muestras de mala calidad o nunca converge. También menciona que las divergencias con las cuales las GAN típicamente minimizan son potencialmente no continuas con respecto a los parámetros del generador, lo que hace que sea difícil de entrenar. proponen en lugar de usar divergencia *KL* usar la métrica de distancia entre probabilidades Wasserstein (también llamada Earth-Mover).

GANs son una clase de modelos generativos que formula en proceso de entrenamiento como un juego, minmax game entre dos redes neuronales: una red generador, que genera datos sintéticos y otra discriminador, que tiene la tarea de diferenciar datos falsos de datos reales del dataset. Wasserstein GAN, hace que de la distancia Wasserstein para calcular el valor la función costo, la cual tiene mejores propiedades teóricas que la gan ordinaria. wGAN requiere que el discriminador (en este contexto algunas veces se le llama criterio) debe estar dentro del espacio de funciones *1-Lipschitz*, (Gulrajani et al., 2017) además proponen una penalización al gradiente *GP*, donde P_r es la distribución de los datos, P_g es el modelo de distribución definido como $\hat{x} = G(\mathbf{z})$, $p(\mathbf{z})$ el vector \mathbf{z} es muestreado desde una distribución simple p como la distribución uniforme o una distribución gaussiana.

Si el discriminador es entrenado optimamente antes del generador, entonces en wGANs argumenta las divergencias de GANs que típicamente minimizan son potencialmente no continuas respecto a los parámetros del generador, **earthmover** también llamada *Wasserstein-1*, distancia $W(q, p)$ informalmente definida como el costo mínimo de transportar masa con el fin de transformar la distribución q en la distribución p (donde el costo es la masa multiplicado por la distancia de transporte).

La ejecución del entrenamiento fue utilizando la GPU, una GTX 970, la memoria de la GPU no fue suficiente para albergar todos los datos, entonces cada época de entrenamiento fue dividida en 30 iteraciones incluyendo 100 imágenes en cada iteración, la duración de cada experimento en el entrenamiento fue de aproximadamente 45 minutos.

5.4 Programación

En un principio se planteo el desarrollo con librería de aprendizaje profundo *Tensorflow* pero en desarrollo se decidió cambiar a *pytorch*, por algunos beneficios, es una librería/framework para aprendizaje profundo fácil de instalar con una API del más fiel al estilo de programación de python, en tabla 3.1 se muestran diferencias entre ambas librerías consideradas para el desarrollo.

Tabla 3.1: Tabla de librerías

Librería	Empresa	Programación	Enfoque	CPU/GPU	Versión
Pytorch	Facebook	Iperativa	Investigación	Sí	1.6
Tensorflow	Google	Declarativa	Empresarial	Sí	0.9

5.5 Evaluación y pruebas

En la sección de resultados los experimentos son divididos en dos enfoques, cualitativo y cuantitativo. En los cualitativos se evalúa a criterio humano, es decir si las muestras generar realmente se parecen a los datos reales en el dataset, la métrica es a criterio solo con la vista. El enfoque cualitativo utiliza una métrica medible que nos arroge un numero, usando un modelo pre-entrenado de clasificación como el *inception-v3* el cual está disponible en para descarga libre, al re-entrenar el ultimo *inception-module* se puede evaluar nuestro modelo de forma cualitativa con otro modelo que no fue utilizado para entrenar el generador. El uso del modelo *inception* es por dos razones, acepta entradas de diferente dimension, y no es costoso re-entrenar el ultimo *inception-module*.

Capítulo 4

Resultados

Después de entrenar el modelo los métodos para calificar el modelo pueden ser cualitativos y cuantitativos. En los métodos cualitativos se involucra la persecución humanas, los medios de validación con una persona que califique con solo ver las imágenes y decidir cual imagen es real y cual es falsa, a esto se le llama prueba de turing en imágenes generadas.

En referente a lo anterior, los modelos generativos no cuentan con una función costo capaz de guiar el entrenamiento hacia un mínimo, es decir no podemos asegurar que un comportamiento en zig-zag sea correcto. recientemente se desarrollo una métrica para los modelos generativos.

Los modelos generativos no pueden ser evaluados con métricas como: porcentaje de instancias correctamente clasificados, curva ROC, TOP-1 o TOP-5, etc. Estos son métodos *supervisados* y no cuentan con una etiqueta a predecir y calcular su rendimiento. Sin embargo recientes investigaciones se han desarrollado métricas como “inception score” y “AM score” para medir de alguna forma el desempeño del modelo de forma cualitativa. ver ([Barratt and Sharma, 2018](#)), ([Zhou et al., 2017](#)),([Salimans et al., 2016](#)).

A continuación se presentan los resultados cuantitativos y cualitativos, los resultados cuantitativos son un conjunto de imágenes sintéticas, con el propósito de apreciar la calidad de las imágenes y su correspondencia con los parámetros de deposición.

1 Experimentos

Las pruebas y experimentos fueron programados y probados con las siguientes características de equipo de cómputo: procesador intel core i7 4710HQ, RAM 16 GB DDR3, sistema operativo gnu/linux ubuntu 17.10, desarrollado con el siguiente software; Lenguaje de programación Python 3.6 mediante la distribución de anaconda 5, librería de aprendizaje profundo *PyTorch*, kit de desarrollo de Nvidia *CUDA* 9.1, librería de optimización en multiplicación de matrices *CudNN* 7.5 para ver más *Aprendice2??*.

Entrenamiento adversario que no converge Antes de presentar los resultados finales, a modo de comparar diferentes resultados con diferentes modelos, en la figura 4.1 se muestran los resultados cuando el entrenamiento no converge adecuadamente. Lo anterior es consecuencia de los parámetros de entrenamiento, una correcta selección de parámetros ayuda a entrenamiento más rápido, mejores resultados, evita errores de colapso. Los resultados obtenidos en la figura 4.1 fueron obtenidos con los siguientes valores de parámetros y técnicas ; λ k = épocas = 100.

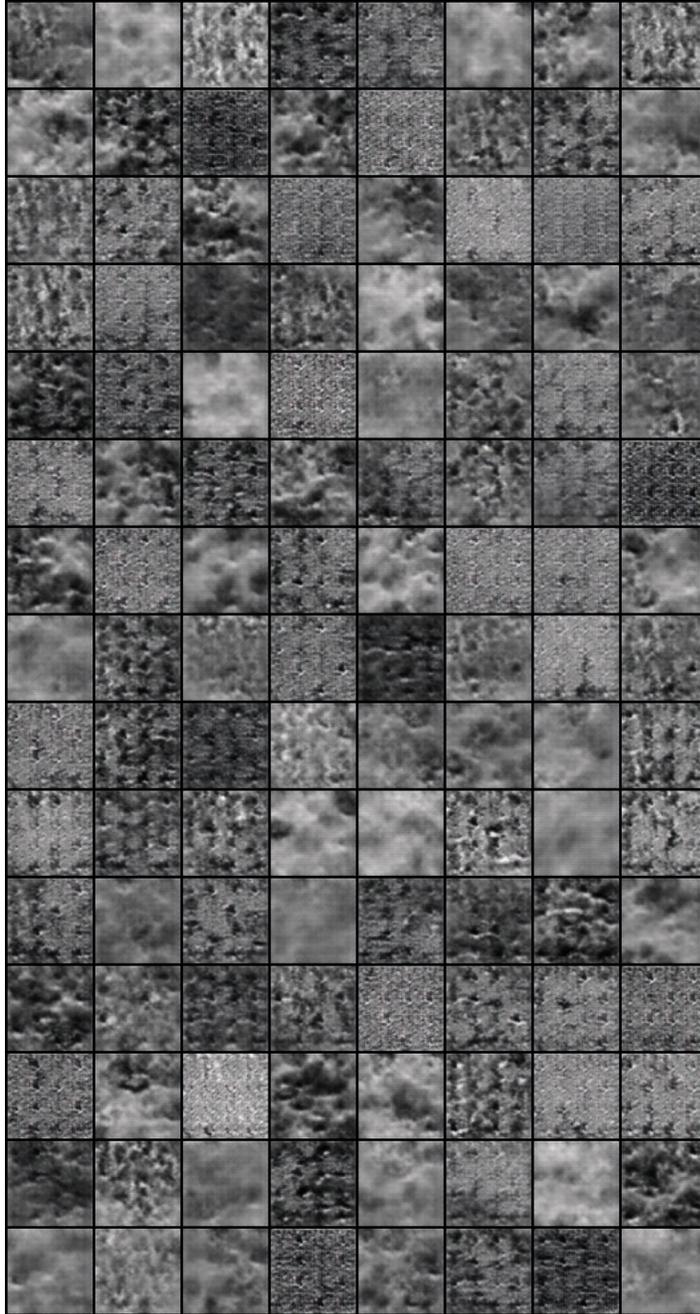


Figura 4.1: Muestra de imágenes que no convergen

Los resultados obtenidos con modelos generacional auto encoders variacionales VAE En la figura 4.2 se muestran imágenes generadas con modelos con un enfoque varacional explicado en la sección 2.11, los resultados en comparación con

GANs son imágenes borrosas, con el mismo método, es resultado de que los VAEs no son capaces de aprender detalles de las imágenes, en (Kingma and Welling, 2013) explica porque la reconstrucción es llevada a cabo por parámetros de distribución σ y μ para múltiples modelos gaussianos.

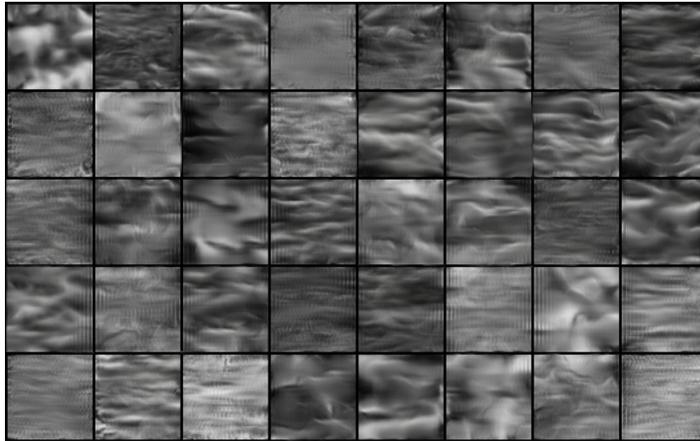


Figura 4.2: Muestras de generación con VAE

Como resultado de este experimento las muestras generadas no son de buena calidad, son borrosas y con pocos detalles, además no es capaz de generar diferentes tipos de muestras que se encuentran en el dataset, este modelo se utilizó principalmente para contrastar resultados con GANs.

1.1 Resultados cualitativos

Como primeros resultados se muestran imágenes generadas en orden de época ascendente sin variables condicionadas, generación libre a base de la distribución de datos en el dataset.

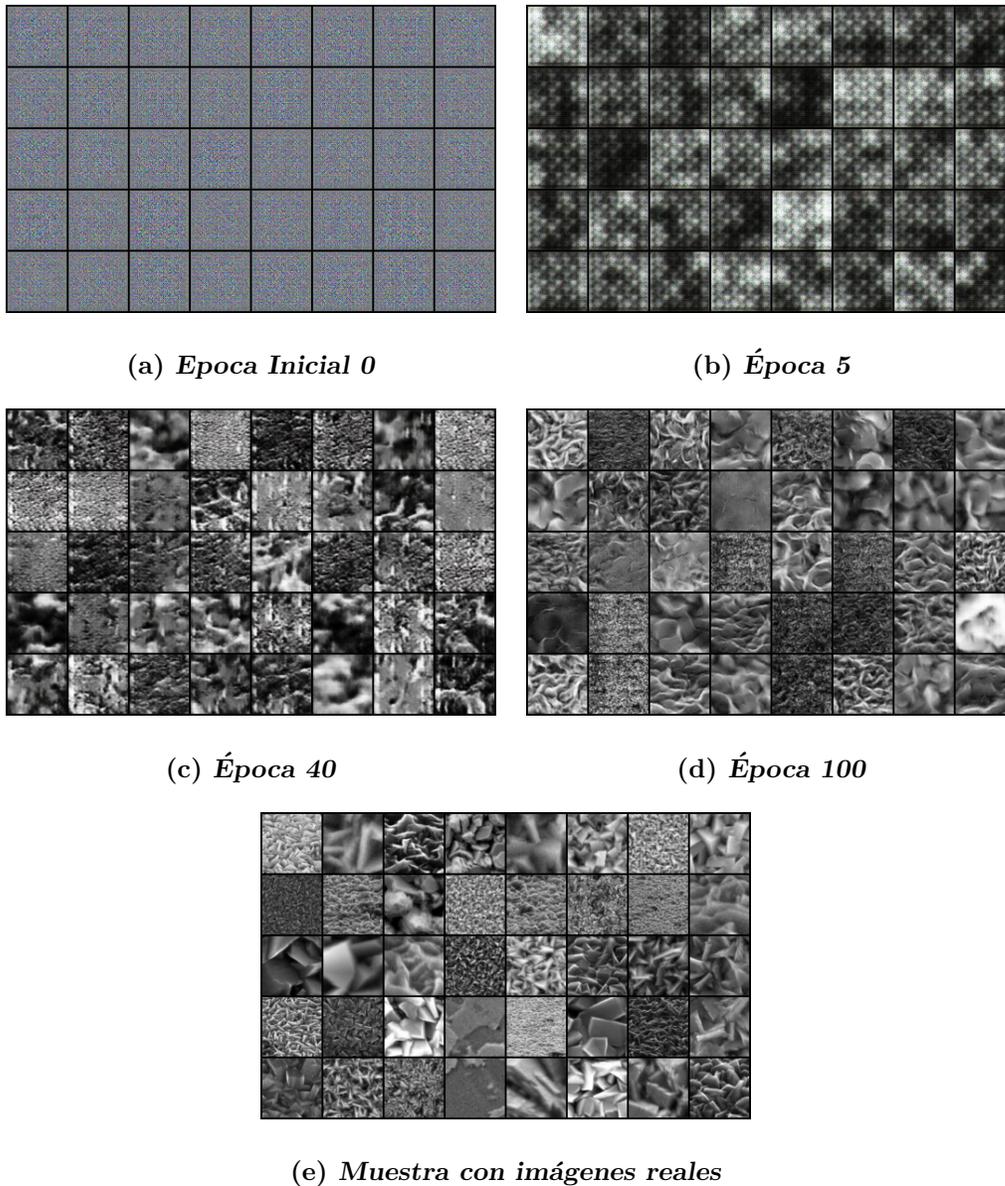


Figura 4.3: Visualización de entrenamiento

Como se aprecie en la figura 4.3(a), las imágenes empiezan formándose a partir de ruido gaussiano, y solo se aprecia eso “ruido” en 4.3(b) el modelo empieza a aprender la relación de colores y formas abstractas de la distribución de los datos, en 4.3(c) las morfologías empiezan a tener forma pero borrosas en 4.3(d). a partir de la época 100 los resultados son mas definidos y se pueden apreciar mejor los detalles en las imágenes a pesar de su baja resolución, las imágenes generadas con producto de un dataset que

luce como Figura (4.4) o (4.5).

2 Generación condicionada

Los parámetros de obtención de imágenes; Magnificación, voltage, emission, distancia. Son parámetros de las leyendas contenidas dentro de las imágenes TEM. Los Experimentos con generación condicionada La generación de imágenes condicionada a variables c discretas. La mayor parte de las imágenes pueden ser codificadas en variables booleanas. (Shen et al., 2017) dice son más fáciles de condicionar por su rango discreto 0,1-falso, verdadero, Magnificación $5000x$ y $10000x$, se codifica como variable booleana verdadero cuando la magnificación igual a $10000x$ de lo contrario falso.

La manera de implementar es por medio de una matriz escasa “*sparse matrices*”, en ella se codifican las variables c . El conjunto variables c analizadas fueron obtenidas de las leyendas de cada imagen, después almacenadas en un archivo de formato `.csv`, con la siguiente información; nombre de la imagen, magnificación, distancia, energía. Posteriormente a la hora de entrenar, dichos archivos son cargados en Python de forma `numpyarray` y son concatenados al vector oculto z , tanto en el generador como en el discriminador. Recordar que el discriminador no es informado directamente del origen de la muestra.

Magnificación A continuación se muestra un entrenamiento con variable condicional de magnificación, la magnificación es la relación entre el tamaño de la imagen y el tamaño del objeto. Es el nivel de acercamiento que se realiza expresada en x veces, es decir un acercamiento de $1000x$ cerca la obtención de la imagen 1000 veces los resultados se muestran el la figura 4.4, en las imágenes.

Distancia Resultados al modificar la variable condicional distancia, d , es la distancia focal de el lente al sustrato que contiene la película delgada.

Energía Resultados al modificar la variable condicional potencia, es la cantidad de potencia que fue utilizada en el SEM para realizar la adquisición de la imagen.

Se lograron los siguientes resultados agrupando y por variación de parámetros. A continuación de muestran gráficas de comportamiento durante el entrenamiento, gráficas e imágenes generadas, los resultados son mostrados por renglones que donde la variable condicional es fija.

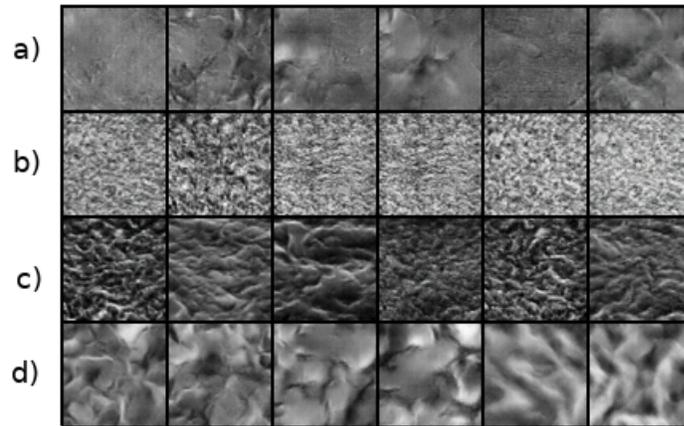


Figura 4.4: Generaciones condicionadas I

En la figura 4.4 Muestra generación con variable; a) Distancia $1000nm$ b) Magnificador a $5000x$ c) Potencia $10.0kV$ d) Magnificador $5000x$ y potencia $15.0kV$.

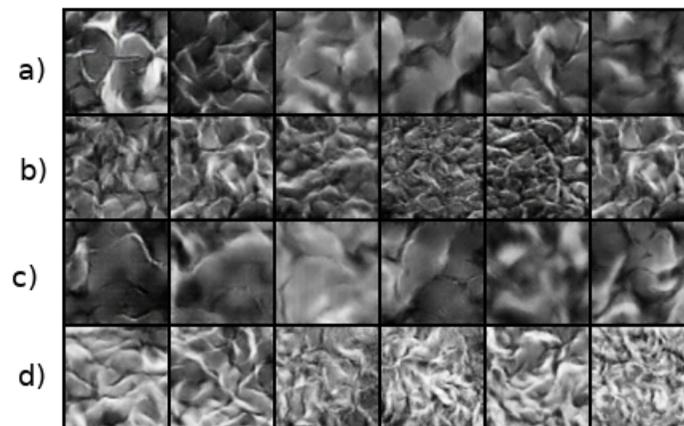


Figura 4.5: Generaciones condicionadas II

En la figura 4.5 muestra; a) Magnificador $10000x$ b) Magnificador $8000x$ c) Distancia $500nm$ d) Potencia $10.0kV$, Magnificador $10000x$.

Las generaciones mostradas son imágenes

3 Resultados cuantitativos

Para obtener resultados cuantitativos que nos ayuden a describir el rendimiento del modelo, fue mediante la métrica “*inception score*” y “*AM score*”. *Inception score* esta métrica se inventó en (Salimans et al., 2016) y se basa en la heurística de que las muestras realistas generadas por la red generativa deben tener un vector de predicción de maxmax de baja entropía cuando pasan a través de una red preentrenamiento, como *Inception* en Imagenet. Una baja entropía para una sola muestra indica que la red reconoce esta imagen como una clase particular de objeto. Sin embargo, el realismo es solo un aspecto del buen desempeño de un modelo generativo. esta métrica a mostrado tener correlación positiva cuando un humano califica el realismo de las imágenes generadas

Otro aspecto muy importante es la cobertura, es decir, cuán diversas son las muestras generadas. Esto sirve como una medida aproximada de si se produjo algún fenómeno de colapso de modo durante el entrenamiento, lo que da como resultado muestras con muy poca variación entre ellas. (falla común en el entrenamiento GAN). La varianza de las muestras se mide por alta entropía sobre las muestras generadas, nos dice que tanto se parecen (existe un correlación positiva) las imágenes generadas contra las imágenes reales en una forma de clasificación.

3.1 Análisis de las tablas

El análisis de los resultados muestran instancias generadas de G con libre generación sin variables condicionales al modelo. La tabla 4.1, variable condicional, precisión del modelo clarificando solo imágenes generadas. En la siguiente figura 4.1, se muestra una evaluación con inception score (sección 3) con modelo clasificador inception-v3.

Tabla 4.1: Tabla generación condicionada

Variable c	Precisión
Magnificación	0.9324
Distancia	0.8202
Energía	0.8962

3.2 Análisis del entrenamiento

Entrenamiento El tiempo de entrenamiento en promedio de al rededor 45 minutos usando la gpu, tambien ayudó Pytorch, drivers y SDK de nvidia actualizados.

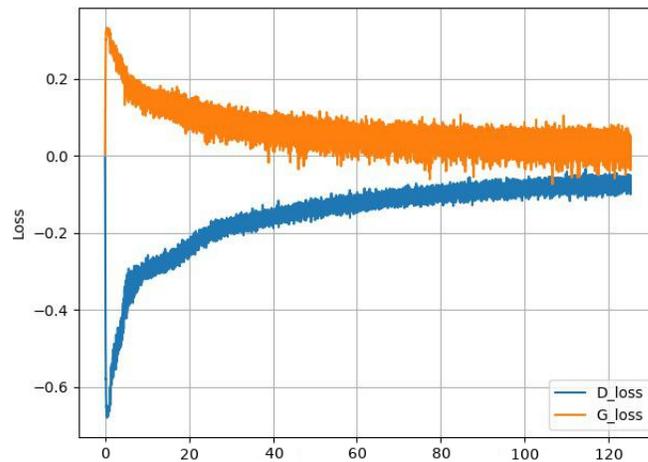


Figura 4.6: Progreso de entrenamiento con 120 épocas

Analizando el comportamiento del valor costo de cada red en figura 4.6, el costo del generador *Gloss* tiene un decremento mientras van ocurriendo las épocas, el valor a partir de la época 100 oscila entre -0.002 y 0.08, considerando que el valor perfecto es 0, dichos valores son estupendos, el costo del discriminador *Dloss* casi toca el 0, recordar que un discriminador perfecto en ~ 0.5 .

Capítulo 5

Conclusiones y trabajo futuro

1 Conclusiones

En presente trabajo de tesis, se abordó el problema de análisis morfológico de nanoestructuras mediante una técnicas de aprendizaje profundo, el cual ha tenido grandes resultados en diferentes áreas dentro del campo de visión artificial. La arquitectura del modelo fue planteada con diferentes técnicas de aprendizaje profundo que permitieron mejorar los resultados, como son redes convolucionales, aprendizaje por transferencia, redes auto-codificadas, entrenamiento adversario.

Las redes generativas adversarias son una nueva forma de generar muestras de una distribución de datos. desde su aparición en 2014 son un punto de inflexión para el modelado generativo y entrenamiento adversario. Han demostrado tener gran desempeño en diferentes aplicaciones que las hacen superiores a otros modelos similares por ejemplo autoencoders variacionales (*VAE*).

A pesar de que *GANs* han tenido gran éxito, el proceso de entrenamiento es lento e inestable, la implementación se tiene que hacer con varios “trucos.”^a tomar en cuenta y no caer en errores de entrenamiento como son; resultados muy borrosos, mode colapse y no convergencia [4.1](#).

Los resultados presentados muestran un gran potencial para la predicción de imágenes *SEM* de soluciones a base de *ZnO* en forma de película delgada, es útil para el proceso de caracterización de propiedades físicas lo que conlleva a la investigación de nuevos nanomateriales que permitan fabricar nuevos y más eficientes dispositivos optoelectrónicos.

A pesar del progreso, evaluar la realidad de los modelos generativos sigue siendo en gran medida un problema abierto y difícil. Sería ideal simplemente tener una función objetiva que mide cuánta realidad asignaría un ser humano a una imagen generada y maximizarla directamente. Sin embargo, esta área está recibiendo mucha atención, por lo que podríamos converger a una respuesta pronto.

2 Trabajo futuro

Se proponen varias mejoras en diferentes puntos de vista. Se propone aumentar el conjunto de datos imágenes para impactar en los resultados. Incluir el análisis para otros óxidos transparentes conductivos como pueden ser *ITO*. También es posible crear un conjunto de imágenes con más meta datos e información de deposición, es decir más con parámetros de deposición físicos se podrá generar imágenes sintéticas más precisas.

Generar imágenes más realistas de alta definición utilizando técnicas como “hiper resolución” por (van den Oord et al., 2016) o con modelos que permitan puedan entrenar de manera estable resoluciones más grandes ambas opciones también se requiere más poder de computo.

Entrenar a partir de nuevos modelos y arquitecturas del estado del arte (Zhao et al., 2017), (Zhai et al., 2016), (Berthelot et al., 2017), (Jaiswal et al., 2017), (Tolstikhin et al., 2017) lo que nos puede dar mejores resultados en los experimentos.

Realizar un análisis de vectores latentes condicionales, puede ser útil para investigar como cambia la morfología con diferentes parámetros de deposición, visualizar un efecto de *morphing* (Chen et al., 2017) sobre las imágenes. con esto es posible realizar operaciones sobre los vectores y descubrir nuevos resultados. Una muestra del potencial de esta área de oportunidades dentro de esta línea de investigación, simulación de proceso de deposición de nano-estructuras, o en microscopía.

Apéndice A

Código/Librerías

1 Código

El repositorio de código fuente desarrollado <https://github.com/Karimx>

2 Librerías

Tensorflow framework de google inc. para el aprendizaje profundo utilizando cpu o gpu. <https://www.tensorflow.org/>.

Compilaciones optimizadas de tensorlow <https://github.com/mind/wheels>

Cuda Entorno de desarrollo para tarjetas de video *GPU* propietario de Nvidia. <https://developer.nvidia.com/cuda-downloads>

CudNN libreria de primitivas para redes neuronales artificiales con aceleracion por *GPU* <https://developer.nvidia.com/cudnn>

MKL libreria de intel Math Kernel Library: Optimización en operaciones de aprendizaje profundo en paralelo <https://software.intel.com/en-us/mkl>.

Anaconda Distribución de python 3 con varios paquetes incluidos. <https://conda.io/docs/download.html>.

Ubuntu Distribución de GNU/linux basada en debian <https://www.ubuntu.com/>.

Pandas librería de python para la gestión de archivos de texto y *datasets* <https://pandas.pydata.org>.

Numpy librería de python para computo numerico y científico <http://www.numpy.org/>.

Sciklearn librería de python para el aprendizaje automático <http://scikit-learn.org/stable/>.

Jupyter Notebook Editor de código fuente tipo notebook interactivo, que soporta múltiples lenguajes de programación. <https://jupyter.org/>.

OpenCV Framework para el procesamiento de imágenes digitales <https://opencv.org/>.

PyTorch framework de [facebook.com](https://www.facebook.com) para el aprendizaje profundo con filosofía de python www.pytorch.com

3 Lista de modelos pre-entrenados

<https://github.com/fchollet/deep-learning-models.git> <https://github.com/tensorflow/models/tree/master/research/slim/nets/nasnet>

Bibliografía

- Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein GAN. 2017. ISSN 1938-7228. URL <http://arxiv.org/abs/1701.07875>. 25
- Pierre Baldi. Autoencoders, Unsupervised Learning, and Deep Architectures. *ICML Unsupervised and Transfer Learning*, pages 37–50, 2012. ISSN 0899-7667. doi: 10.1561/22000000006. 23
- Shane Barratt and Rishi Sharma. A Note on the Inception Score. 2018. URL <http://arxiv.org/abs/1801.01973>. 45
- David Berthelot, Thomas Schumm, and Luke Metz. BEGAN: Boundary Equilibrium Generative Adversarial Networks. pages 1–10, 2017. doi: 1703.10717. URL <http://arxiv.org/abs/1703.10717>. 55
- Matthias Bethge. Note on the evaluation of generative models. pages 1–10, 2016. 30
- R R Biswal, A Maldonado, and M De L Olvera. Electrical and Optical Properties of In-doped ZnO Thin Films via Ultrasonic Spray Pyrolysis. pages 407–410, 2013. 14
- M Bremer. Multiple Linear Regression. *Handbook of Regression Analysis*, (1):1–21, 2013. ISSN 1633-4760. doi: 10.1002/9781118532843.ch1. URL <http://doi.wiley.com/10.1002/9781118532843.ch1>. 20
- Xi Chen, Yan Duan, Rein Houthoofd, John Schulman, Ilya Sutskever, and Pieter Abbeel. InfoGAN : Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets. 2017. 27, 55
- Nader Ebrahimi, Mahmoud Shehadeh, and Kristin McCullough. Bayesian Analysis for Robust Synthesis of Nanostructures. *Nanoscale Systems: Mathematical Modeling, Theory and Applications*, 1:172–186, 2012. ISSN 2299-3290. doi: 10.2478/nsmmt-2012-0010. URL <http://www.degruyter.com/view/j/nsmmt.2012.1.issue/nsmmt-2012-0010/nsmmt-2012-0010.xml>. 31

- Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. 521(7553):436–444. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature14539. URL <http://www.nature.com/articles/nature14539>. bibtex: lecun_deep_2015. 21
- Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey. Adversarial Autoencoders. 2015. URL <http://arxiv.org/abs/1511.05644>. 27
- Itzik Malkiel, Achiya Nagler, Michael Mrejen, Uri Arieli, Lior Wolf, and Haim Suchowski. Deep Learning for Design and Retrieval of Nano-photonic Structures. *arXiv*, (DI), 2017. URL <https://arxiv.org/ftp/arxiv/papers/1702/1702.07949.pdf><http://arxiv.org/abs/1702.07949>. 31
- Mehdi Mirza and Simon Osindero. Conditional Generative Adversarial Nets. pages 1–7, 2014. ISSN 10495258. URL <http://arxiv.org/abs/1411.1784>. 27
- Mustafa Mustafa, Deborah Bard, Wahid Bhimji, Rami Al-Rfou, and Zarija Lukić. Creating Virtual Universes Using Generative Adversarial Networks. 2017. URL <http://arxiv.org/abs/1706.02390>. 31
- Vu Nguyen, Tomas F Yago Vicente, Maozheng Zhao, Minh Hoai, and Dimitris Samaras. Shadow Detection with Conditional Generative Adversarial Networks. *International Conference on Computer Vision*, 2017. doi: 10.1109/ICCV.2017.483. URL <http://www3.cs.stonybrook.edu/~minhhoai/papers/scGAN{ }ICCV17.pdf>. 31
- Anton Osokin, Anatole Chessel, Rafael E. Carazo Salas, and Federico Vaggi. GANs for Biological Image Synthesis. 2017. doi: 10.1109/ICCV.2017.245. URL <http://arxiv.org/abs/1708.04692>. 31
- David Pfau. Unrolled Generative Adversarial Networks. pages 1–25, 2017. 22, 30
- Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. pages 1–16, 2015. ISSN 0004-6361. doi: 10.1051/0004-6361/201527329. URL <http://arxiv.org/abs/1511.06434>. 24, 29
- S K M Reza and K M Iftekharuddin. Multi-class Abnormal Brain Tissue Segmentation Using Texture Features. 2013:38–42, 2013. 20
- Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved Techniques for Training GANs. pages 1–10, 2016. ISSN 09246495. doi: arXiv:1504.01391. URL <http://arxiv.org/abs/1606.03498>. 43, 45, 52
- Mahmoud Shehadeh, Nader Ebrahimi, and Abel Ochigbo. Predicting the type of nanostructure using data mining techniques and multinomial logistic regression. *Procedia*

-
- Computer Science*, 12:392–397, 2012. ISSN 18770509. doi: 10.1016/j.procs.2012.09.092. 31
- Shiwei Shen, Guoqing Jin, Ke Gao, and Yongdong Zhang. APE-GAN : Adversarial Perturbation Elimination with GAN. 2017. URL <http://arxiv.org/abs/1707.05474>. 50
- Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. pages 1–14, 2014. ISSN 09505849. doi: 10.1016/j.infsof.2008.09.005. URL <http://arxiv.org/abs/1409.1556>. 36
- Changwon Suh, Chris W. Gorrie, John D. Perkins, Peter A. Graf, and Wesley B. Jones. Data-mining-aided mapping of structure-property relationships for combinatorially generated Co-doped ZnO thin films. *Conference Record of the IEEE Photovoltaic Specialists Conference*, pages 2497–2502, 2010. ISSN 01608371. doi: 10.1109/PVSC.2010.5614708. 31
- Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, Chapel Hill, and Ann Arbor. Going Deeper with Convolutions. pages 1–9, 2014. ISSN 10636919. doi: 10.1109/CVPR.2015.7298594. 30
- Ilya Tolstikhin, Sylvain Gelly, Olivier Bousquet, Carl-Johann Simon-Gabriel, and Bernhard Schölkopf. AdaGAN: Boosting Generative Models. pages 1–31, 2017. URL <http://arxiv.org/abs/1701.02386>. 35, 55
- Aaron van den Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. Pixel Recurrent Neural Networks. 2016. URL <http://arxiv.org/abs/1601.06759>. 55
- Aparna Varde, Jianyu Liang, Elke Rundensteiner, and Richard Sisson Jr. Mining Images of Material Nanostructure Data. *Icdcit*, pages 403 – 413, 2006. 30
- Sebastien C. Wong, Adam Gatt, Victor Stamatescu, and Mark D. McDonnell. Understanding Data Augmentation for Classification: When to Warp? *2016 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2016*, 2016. doi: 10.1109/DICTA.2016.7797091. 38
- Shuangfei Zhai, Yu Cheng, Rogerio Feris, and Zhongfei Zhang. Generative Adversarial Networks as Variational Training of Energy Based Models. pages 1–12, 2016. URL <http://arxiv.org/abs/1611.01799>. 27, 55
- Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, and Dimitris Metaxas. StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks. 2016. doi: 10.1109/ICCV.2017.629. URL <http://arxiv.org/abs/1612.04357><http://arxiv.org/abs/1612.03242>. 25
-

Junbo Zhao, Michael Mathieu, and Yann Lecun. Energy - Based Generative and Adversarial Networks. (2014):1–17, 2017. 55

Zhiming Zhou, Weinan Zhang, and Jun Wang. Inception Score, Label Smoothing, Gradient Vanishing and $-\log(D(x))$ Alternative. 2017. URL <http://arxiv.org/abs/1708.01729>. 30, 45