**Instituto Tecnológico**

**y de Estudios Superiores de Occidente**

Reconocimiento de validez oficial de estudios de nivel superior según acuerdo secretarial 15018, publicado en el Diario Oficial de la Federación del 29 de noviembre de 1976.

Departamento de Electrónica, Sistemas e Informática

Maestría en Sistemas Computacionales



**Desarrollo de un módulo de visión computacional para vehículos aéreos no tripulados utilizando Deep Learning**

Trabajo recepcional que para obtener el grado de

Maestro en Sistemas Computacionales

Presenta: José Luis Magaña Vázquez

Jehovany Alejandro Rodríguez Vázquez

Asesor Dr. José Francisco Cervantes Álvarez

Tlaquepaque, Jalisco. Haga clic aquí para escribir una fecha..

[La portada se deberá actualizar con la versión oficial disponible en: <https://www.iteso.mx/titulacionposgrados>]

[SI EL DOCUMENTO ES ESCRITO EN INGLES, se debe colocar la portada principal en inglés, y una segunda portada en español.]

AGRADECIMIENTOS

El autor desea dar las gracias a [Escriba los Agradecimientos].

[Para quienes hayan recibido apoyo CONACYT, se debe incluir aquí el agradecimiento por la Beca recibida, así como el número de beca – no confundir con número de CVU.]

[Se sugiere incluir agradecimiento al ITESO si se les apoyó con descuento o beca. Lo mismo para la empresa donde trabajan. También podrían incluir aquí agradecimiento a sus profesores, asesores o directores del trabajo.]

[SI EL DOCUMENTO ES ESCRITO EN INGLES, la primera página de agradecimientos es en inglés, y la segunda en español.]

DEDICATORIA

El autor dedica esta tesis a [Escriba la Dedicatoria].

[SI EL DOCUMENTO ES ESCRITO EN INGLES, la primera página de dedicatoria va en inglés, y se incluye una seguna en español]

RESUMEN

[Hablar brevemente sobre las partes de este trabajo, y lo que se presenta en cada uno] Se presenta una breve introducción a [al problema principal que resuelve este trabajo], el cual tiene como objetivo principal [mencionar objetivo principal], resolviendo de manera particular [a los objetivos específicos. [Se presenta el desarrollo del trabajo y sus principales resultados.] Finalmente, [se presentan las conclusiones del trabajo]. [Este resumen cuenta con 250 palabras máx.]

[SI EL DOCUMENTO SE ESCRIBE EN INGLÉS, la primera página de resumen va en inglés, y se deberá incluir una segunda página con el resumen en español.]

TABLA DE CONTENIDO

[Dar clic en el menú de Referencias/ Dar clic en Tabla de Contenido/ Seleccionar Tabla de Contenido Personalizada/ Seleccionar Formato Formal, 4 Niveles, Dar Clic en Opciones. En la ventana de Opciones, asignar Nivel de TDC 1 a Título Intro. Dar clic en Aceptar, y de nuevo en Aceptar].

[AGRADECIMIENTOS 3](#_Toc451186816)

[DEDICATORIA 3](#_Toc451186817)

[RESUMEN 3](#_Toc451186818)

[TABLA DE CONTENIDO 3](#_Toc451186819)

[LISTA DE FIGURAS 3](#_Toc451186820)

[LISTA DE TABLAS 3](#_Toc451186821)

[LISTA DE ACRÓNIMOS Y ABREVIATURAS 3](#_Toc451186822)

[1. INTRODUCCIÓN 3](#_Toc451186823)

[1.1. Antecedentes 3](#_Toc451186824)

[1.2. Justificación 3](#_Toc451186825)

[1.3. Problema 3](#_Toc451186826)

[1.4. Hipótesis 3](#_Toc451186827)

[1.5. Objetivos 3](#_Toc451186828)

[1.5.1. Objetivo General: 3](#_Toc451186829)

[1.5.2. Objetivos Específicos: 3](#_Toc451186830)

[1.6. Novedad científica, tecnológica o aportación 3](#_Toc451186831)

[2. ESTADO DEL ARTE o de la TÉCNICA 3](#_Toc451186832)

[2.1. Tema relacionado 1 3](#_Toc451186833)

[2.2. Tema relacionado 2 3](#_Toc451186834)

[3. MARCO TEÓRICO/CONCEPTUAL 3](#_Toc451186835)

[3.1. Concepto básico 1 3](#_Toc451186836)

[3.2. Esquema básico 2 3](#_Toc451186837)

[3.2.1. Si se requiere subtema 3](#_Toc451186838)

[3.2.1.1. Subsubtema 3](#_Toc451186839)

[4. DESARROLLO METODOLÓGICO 3](#_Toc451186840)

[4.1. Levantamiento de requerimientos 3](#_Toc451186841)

[5. RESULTADOS Y DISCUSIÓN 3](#_Toc451186842)

[5.1. Resultados 3](#_Toc451186843)

[5.2. Discusión 3](#_Toc451186844)

[6. CONCLUSIONES 3](#_Toc451186845)

[6.1. Conclusiones 3](#_Toc451186846)

[6.2. Trabajo Futuro 3](#_Toc451186847)

[BIBLIOGRAFÍA 3](#_Toc451186848)

LISTA DE FIGURAS

[Figura 1. Ventana que se abre para insertar una imagen. 3](#_Toc451189394)

[En el menú de Referencias, seleccionar Insertar Tabla de Ilustraciones. Si se desea insertar para Figuras, solo se debe seleccionar el rótulo de Figura.]

LISTA DE TABLAS

[Tabla 1. Presentación de información en tablas. 3](#_Toc451189447)

[En el menú de Referencias, seleccionar Insertar Tabla de Ilustraciones. Si se desea insertar para Tablas, solo se debe seleccionar el rótulo de Tabla.]

LISTA DE ACRÓNIMOS Y ABREVIATURAS

[Del lado izquierdo va el acrónimo o abreviatura y del lado derecho su significado. La columna de en medio se deja en blanco. Al final, se deben quitar los bordes de la tabla]

|  |  |  |
| --- | --- | --- |
| **Abreviación** | **Significado** | **Referencia** |
| VGG | Visual Geometry Group | Hablar de una red neuronal VGG es hablar de una red neuronal convolucional profunda para reconocimiento de objetos desarrollada y entrenada por el reconocido grupo de geometría visual de Oxford (VGG por sus siglas en ingles) el cual logró un muy buen rendimiento en el conjunto de datos ImageNet. |
| ILSVRC | ImageNet Large Scale Visual Recognition Challenge | ILSVRC es una competencia anual de visión computacional desarrollada sobre un subconjunto de datos de visión computacional disponible públicamente llamado ImageNet |
| CIFAR | Canadian Institute For Advanced Research | Un conjunto de datos CIFAR-10 es una colección de imágenes que comúnmente se usan para entrenar algoritmos de aprendizaje automático y visión computacional. |
| COCO | Common Objects in Context | COCO es una base de datos que tiene como objetivo permitir futuras investigaciones para la detección de objetos, segmentación de instancias, subtitulado de imágenes y localización de puntos clave de personas. |
| SDG | Stochastic Gradient Descent | El descenso de gradiente estocástico es un método para optimizar una función objetivo con propiedades de suavidad adecuadas. Puede considerarse como una aproximación estocástica de la optimización del descenso del gradiente, ya que reemplaza el gradiente real por una estimación del mismo. |
| VLAD | Vector of Locally Aggregated Descriptors | Los descriptores agregados localmente (VLAD) es un tipo de método eficiente de codificación de súper vector. El método de codificación VLAD se utiliza para transformar las características locales en representaciones vectoriales de tamaño fijo. La clasificación se utiliza para evaluar los resultados finales de la clasificación. |
| PDEs | Partial Differential Equations | En matemáticas, una ecuación diferencial parcial es una ecuación diferencial que contiene funciones multivariables desconocidas y sus derivadas parciales. Las PDE se usan para formular problemas que involucran funciones de varias variables, y se resuelven a mano o se usan para crear un modelo de computadora. |
| YOLO | You Only Look Once | YOLO es una inteligente red neuronal convolucional (CNN) para detectar objetos en tiempo real. El algoritmo aplica una única red neuronal a la imagen complete, para después dividirla en regiones y predecir cuadros delimitadores así como las probabilidades para cada región. |
| mAP | mean Average Precision | La precisión promedio es una medida que combina la recuperación y la precisión para resultados de recuperación clasificados. Para una necesidad de información, la precisión promedio es la media de los puntajes de precisión después de recuperar cada documento relevante. |
| R-CNN | Region Convolutional Neural Network | Una red neuronal convolucional (CNN) es principalmente para clasificación de imágenes, mientras que una red neuronal convolucional regional (RCNN) se usa para detección de objetos. Una CNN solo te dice la clase de los objetos pero no donde están localizados. |
| DPM | Deformable Part Models | Es un núcleo FPGA IP de detección de objetos basada en el aprendizaje, desarrollada para aplicaciones de visión integrada. Utiliza un modelo basado en partes estructurado en estrella, definido por un filtro raíz mas un conjunto de filtros de partes y modelos de deformación asociados. |
|  |  | |

# INTRODUCCIÓN

**Resumen:** En este capítulo se presenta brevemente los antecedentes del objeto de estudio, justificación del objeto de estudio, justificación y la definición del problema. [hipótesis si la investigación lo justifica].

La IA en etapas tempranas se dedicó a resolver problemas que son intelectualmente complicados para los humanos, aunque relativamente fácil para una computadora. Problemas que pueden ser resultado con una serie de instrucciones en lenguaje formal. El reto de la IA ha sido realzar tareas que son fáciles de realizar por los humanos, aunque difícil de describir en un lenguaje formal. Problemas o tareas que parecen instintivas como el reconocimiento de voz o el reconocimiento de rostros y estados de ánimo.

Una forma de solución para la complejidad de definir una serie de instrucciones en lenguaje formal es que las computadoras puedan aprender a través de la experiencia y una jerarquía de conceptos más simples para generar conceptos más complejos. De esta forma de aprender, se evita que el humano defina en lenguaje formal todo el conocimiento necesario para la computadora.

Si pudiéramos ver cómo estos conceptos más complejos se forman a través de otros más simples, podríamos verlo como una profundidad de varias capas de conceptos. Este tipo de aproximación a la inteligencia artificial se le conoce como” Deep Learning” o aprendizaje profundo.

En el uso de aprendizaje profundo, se requieren redes neuronales. Las redes neuronales son un conjunto de algoritmos que son diseñados para reconocer patrones. Por ello el interés de utilizarlos para un módulo de visión. Estas redes interpretan datos de entrada a través de etiquetar, agrupar y clasificar. Los patrones que reconocen son numéricos, contenidos en vectores que representan cualquier dato del mundo real o del problema a tratar, como sonido, texto o en este caso imágenes. Estos algoritmos ayudan a agrupar datos sin etiquetar de acuerdo a las similitudes con las entradas de ejemplo y clasifican datos, así como pueden extraer características para obtener un conjunto de las mismas para la representación de los datos.

En el caso del procesamiento de imágenes se usarán un tipo de red neuronal llamada red neuronal convolucional o CNN por sus siglas en inglés, donde el campo receptivo de una unidad convolucional con un vector de pesos asignados (filtro) es desplazado a través de un arreglo de 2-dimensiones de la entrada, como los pixeles de una imagen de entrada, dando como resultado un arreglo de 2-dimensiones el cual puede proveer como entrada para subsecuentes capas de la red neuronal.[1, p. 10]. Este tipo de red neuronal ha estado presente en varios algoritmos que han ganado competencias como IMAGENET Large

## Antecedentes

[incluir desarrollos, trabajos, resultados, publicaciones, etc., similares al objeto de este trabajo. Mencionar brevemente, sin entrar en el detalle de los mismos. El detalle se realizará en la sección de Estado del Arte o de la Técnica]

## Justificación

[Razones científicas, económicas, y/o sociales que muestren la necesidad de este trabajo]

## Problema

[Describir problema práctico, definir problema científico]

## Hipótesis

[Si la investigación lo justifica]

## Objetivos

#### Objetivo General:

[Definir de ser necesario. No incluir métodos o procedimientos]

#### Objetivos Específicos:

[Definir, no incluir métodos o procedimientos, numerados]

## Novedad científica, tecnológica o aportación

[Especificar lo novedoso de este trabajo o la aportación que brinda a la empresa o usuario]

# ESTADO DEL ARTE o de la TÉCNICA

**Resumen:** En este capítulo se presenta un resumen de los trabajos relacionados con Deep Learning.

## Aprendizaje Profundo Residual para Reconocimiento de Imágenes (Deep Residual Learning for Image Recognition).

|  |  |  |
| --- | --- | --- |
| Problema | ¿Aprender mejores redes es tan fácil como apilar más capas? | |
| Desarrollo | Descripción de factores | Cuando redes profundas son capaces de empezar a converger, un problema de degradación ha sido expuesto: con la profundidad de la red aumentando, la precisión se satura y se degrada rápidamente. Inesperadamente dicha degradación no esta causada por el sobreajuste, y agregar más capas a un modelo adecuadamente profundo conduce a un error de entrenamiento más alto. |
| Conclusiones en relación con la. Investigación | Las redes de alta velocidad no han demostrado ganancias de precisión con profundidades extremadamente aumentadas (arriba de 100 capas) |
| Bibliografía | He2015 | |

## YOLO (Solo miras una vez): Detección unificada de objetos en tiempo real (You Only Look Once: Unified, Real-Time Object Detection)

|  |  |  |
| --- | --- | --- |
| Problema | YOLO como algoritmo para identificación de imágenes tomadas por vehículos aéreos no transportados (UAV) | |
| Desarrollo | Descripción de factores | Limitaciones de YOLO   * En Yolo existen restricciones espaciales para las predicciones dentro de los cuadros delimitadores, solo pueden tener una clase (Árbol, Surcos, Caminos, Construcciones, por mencionar algunos) * El algoritmo se complica al tratar de predecir los objetos pequeños. |
| Conclusiones en relación con la. Investigación | YOLO comete (por mucho) menos errores en segundo plano que Fast R-CNN (Otro sistema de detección), sin embargo comete mas errores en la detección de objetos pequeños en comparación con otros modelos, pero es fácil de entrenar. |
| Bibliografía | Redmon2016 | |
|  |  | |

A screenshot of a cell phone

Description automatically generated

## Abreviaciones

# MARCO TEÓRICO/CONCEPTUAL

**Resumen:** En este capítulo se presentan las bases teóricas y conceptuales sobre redes neuronales y visión.

## Aprendiendo a partir de datos

Existen varios acercamientos de aprendizaje máquina automático, pero uno de los acercamientos más exitosos ha sido el llamado aprendizaje basado en el gradiente. El modelo de aprendizaje calcula una función [3]donde es la p-th patrón de entrada, representa la colección de parámetros ajustables en el sistema y puede ser interpretada como la etiqueta o clase del patrón , o la probabilidad asociada a cada clase a predecir. Es decir que a partir de una entrada se predice la clasificación deseada por medio de una función F.

[3], es una función que mide la discrepancia o error entre el cuál es el valor correcto que deseamos predecir a partir del patrón de entrada y la predicción . Esto implica tener una base de datos con los patrones de entrada con su respectiva salida correcta para poder realizar esta operación con la cuál se entrena el modelo. Con el cálculo del error de cada predicción se puede calcular el promedio y obtener un error de entrenamiento que se llamará función de costo o pérdida. Muchos algoritmos de aprendizaje tratan de minimizar .

## Gradiente Descendiente

El algoritmo de gradiente descendiente intenta minimizar la función de costo a través de variaciones en los pesos W. Estas variaciones son introducidas a cada peso partir del gradiente de la función de perdida con respecto a los pesos. Pero para ello se tiene que cumplir que los parámetros de W son valores reales para los cual la función de costo es una función continua y diferenciable, de esta manera se introducen variaciones en cada parámetro de W iterativamente de la siguiente forma:

Existen variaciones del gradiente descendiente siendo el presentado uno de los más simples.

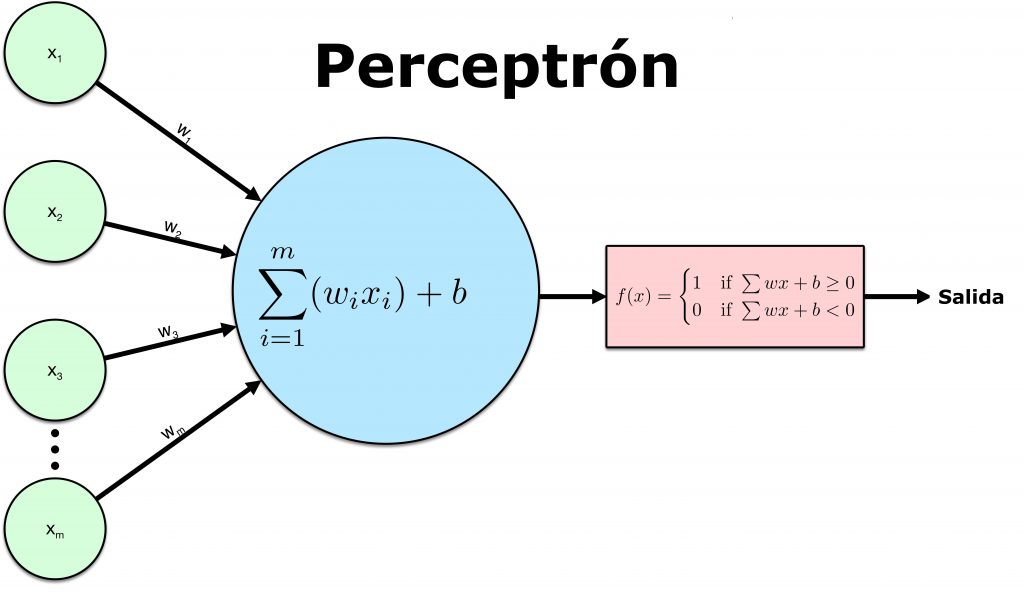
## Perceptrón

El perceptrón es un modelo computacional de una neurona[4], donde una sola neurona tiene varias entradas (dentritas), un cuerpo celular y una salida (axón). Replicando esto el perceptrón tiene varias entradas y una sola salida, el cuál consiste en un vector de “pesos” **w** = [w1 … wm], un peso para cada entrada más un “bias” o tendencia cuyo parámetro es descrito como b.

Con los parámetros dados **w** y b el perceptrón realiza el siguiente cálculo:

Donde se llama función de activación.

El algoritmo del perceptrón busca obtener o aprender los pesos **w** tales que dados los datos de cada ejemplo de entrenamiento , se obtenga la correcta clasificación binaria , donde será 1 o 0 en este caso.



<http://blog.josemarianoalvarez.com/2018/06/10/el-perceptron-como-neurona-artificial/>

Modelo del perceptrón

A continuación, se presenta el algoritmo del perceptrón[4]:

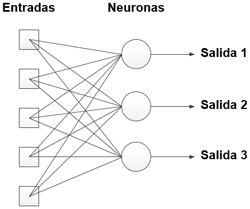
1. Inicializar los parámetros **w** y b
2. Durante N iteraciones (definidas por el desarrollador) o hasta que los pesos no cambien realizar lo siguiente:
   1. Para cada ejemplo de entrenamiento con respuesta
3. Calcular
4. Si continua
5. Si no, actualizar todos los pesos como a continuación

Donde si existe una configuración de pesos **w** con los cuales se puedan clasificar correctamente cada elemento de entrenamiento, el algoritmo lo encontrará, aunque en ejemplos reales no siempre es posible, el algoritmo puede encontrar una configuración de pesos **w** para clasificar correctamente un porcentaje de los elementos de entrenamiento.

En casos donde se necesite más clases se puede realizar incrementando la cantidad de perceptrones por cada clase, siendo el valor 0 el indicador de no pertenencia a la clase y el número 1 la pertenencia a la clase, teniendo así una llamada red neuronal.

## Red Neuronal Clásica

Cómo se ha visto con el perceptrón, una red neuronal clásica consiste en conjunto de modelos computacionales llamadas neuronas, las cuales tienen parámetros ajustables **w** y “bias” que, por medio de un algoritmo, la red puede aprender a clasificar correctamente las clases a partir de los datos de entrada .

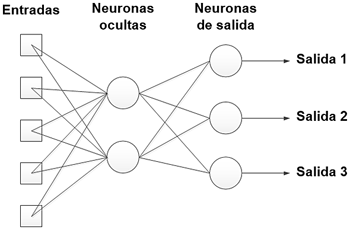


<https://www.ediciones-eni.com/open/mediabook.aspx?idR=6d7e746f6a1fe07834a44f834c84d6ba>

Modelo de una red neuronal

En la imagen se puede observar un modelo de red neuronal, cuyas neuronas tienen una salida. Ahora, las salidas de las neuronas en la imagen podríamos tomarlas como entrada para un siguiente conjunto de neuronas, a cada conjunto de neuronas se les llama capas, siendo una capa las neuronas que reciben la entrada original, una segunda capa de neuronas como entrada la salida de la primera capa, teniendo una llamada red neuronal multicapa, donde a las capas entre la entrada y la salida se les llama capas ocultas.

De esta manera se pueden obtener redes neuronales de n capas, por lo que cada capa aumenta el tamaño o la “profundidad” de la red neuronal, y es aquí de donde procede el termino aprendizaje profundo o también conocido en inglés como “Deep learning”.



<https://www.ediciones-eni.com/open/mediabook.aspx?idR=dd1e7c64bbcc8628f67483605a21d1d2>

## Red Neuronal “Feed Forward” y “Back Propagation”

Son redes neuronales cuyo grafo con dirección no contienen ciclos, normalmente las redes neuronales multicapa utilizan un algoritmo de gradiente descendiente para calcular el valor de modificación de los pesos **w,** donde la diferencia de las neuronas del tipo perceptrón es el cambio de la función de activación, donde la función de activación del perceptrón está dada por:

Esto debido que para hacer uso del gradiente descendiente necesitamos que la función de costo o pérdida sea diferenciable y continua para el conjunto de valores reales de los pesos **w.** Dado que la función de pérdida [3] depende de la función de activación, la función del perceptrón no cumple los requisitos para implementar el gradiente descendiente.

Existen varias funciones de activación, una de las funciones usadas para las redes neuronales “feed forward” es la función sigmoide:

La cual es derivable:

Donde

Por lo cual se puede utilizar como función de activación y poder calcular el gradiente de la función de perdida para actualizar los pesos **w.** A dicho proceso con el cuál se calcula el gradiente a partir de la función de costo o pérdida y se actualizan los pesos **w** se le llama “Back Propagation”.

## Red neuronal convolucional

BIBLIOGRAFÍA

[1] J. Schmidhuber, «Deep Learning in Neural Networks: An Overview», *Neural Networks*, vol. 61, pp. 85-117, ene. 2015, doi: 10.1016/j.neunet.2014.09.003.

[2] «ImageNet Large Scale Visual Recognition Competition (ILSVRC)». http://image-net.org/challenges/LSVRC/ (accedido oct. 07, 2019).

[3] Y. LeCun, L. Bottou, Y. Bengio, y P. Haffner, «Gradient-Based Learning Applied to Doument Recognition», *Proc. of the IEEE*, November 18.

[4] E. Charniak, *Introduction to Deep Learning*, 1st ed. The MIT Press, 2018.

[5] «Red neuronal artificial», *Wikipedia, la enciclopedia libre*. oct. 04, 2019, Accedido: oct. 06, 2019. [En línea]. Disponible en: https://es.wikipedia.org/w/index.php?title=Red\_neuronal\_artificial&oldid=119954405.

[6] V. E. Bondarenko, «Artificial neural networks», *Salem Press Encyclopedia of Science*. Salem Press, 2018.

[7] «t8neuronales.pdf». Accedido: oct. 06, 2019. [En línea]. Disponible en: http://www.sc.ehu.es/ccwbayes/docencia/mmcc/docs/t8neuronales.pdf.

[8] «Dialnet-ENTRENAMIENTODEUNAREDNEURONALARTIFICIALUSANDOELALG-4844874.pdf». .

[9] S. Shalev-Shwartz y S. Ben-David, *Understanding Machine Learning from Theory to Algorithms*. Cambridge University Press, 2015.

**APÉNDICE A. Título**

**APÉNDICE B. Título**