
ESCUELA TÉCNICA SUPERIOR DE
INGENIERÍA DE TELECOMUNICACIÓN

GRADO EN INGENIERÍA EN INGENIERÍA EN
SISTEMAS AUDIOVISUALES Y MULTIMEDIA

**TRABAJO FIN DE GRADO**

# DEEP LEARNING EN SENSORES RGBD

Autor: David Butragueño Palomar
Tutor:

Curso académico 2016/2017

# Prefacio

# Agradecimientos

# Índice general

**1. Introducción** **1**

**2. Infraestructura** **2**
2.1. Caffe . . . . . 2
2.1.1. Línea de comandos . . . . . 2
2.1.2. Python . . . . . 2
2.1.3. Capas de una red neuronal . . . . . 3
2.2. JdeRobot . . . . . 6
2.3. Base de datos LMDB . . . . . 7
2.3.1. LMDB con Python . . . . . 8

**3. Clasificación** **9**
3.1. Red Neuronal . . . . . 9
3.1.1. Estructura de la red . . . . . 9
3.1.2. Definición del solucionador . . . . . 12
3.1.3. Entrenamiento de la red . . . . . 13
3.2. Base de datos MNIST . . . . . 14
3.2.1. Sección de datos . . . . . 14
3.2.2. Sección de etiquetas . . . . . 15

**Bibliografía** **17**

# Índice de figuras

3.1. Estructura red neuronal . . . . . 12
3.2. Fichero PGM . . . . . 15
3.3. Imagen base de datos MNIST . . . . . 15

# Capítulo 1

# Introducción

# Capítulo 2

# Infraestructura

## 2.1. Caffe

Caffe es un framework de aprendizaje profundo desarrollado por Berkeley AI Research (BAIR) y por contribuyentes de la comunidad. Fue creado por Yangqing Jia durante su doctorado en UC Berkeley. Caffe está publicado bajo la licencia BSD 2-Clause.

### 2.1.1. Línea de comandos

Caffe dispone de una interfaz de línea de comandos llamada *cmdcaffe* la cual es la herramienta utilizada por Caffe para el entrenamiento del modelo y el diagnóstico del mismo. Los principales comandos que se pueden ejecutar son:

- **Entrenamiento:** Con el comando *caffe train* es posible aprender modelos desde cero,
- **Test:** El comando *caffe test* puntúa los modelos ejecutándolos en la fase de test e informa de la salida de la red como su puntuación. En primer lugar se informa la puntuación por lotes de datos de entrada y finalmente el promedio general.
- **Comparación:** El comando *caffe time* compara el modelo capa a capa. Esto es util para comprobar el rendimiento del sistema y medir los tiempos de ejecución relativos a los modelos.

### 2.1.2. Python

La interfaz de Python *pycaffe* contiene el módulo de Caffe y sus propios scripts en la ruta caffe/python. Con el comando *import caffe* se importará esta interfaz, pudiendo así cargar diferentes modelos de Caffe, manejar instrucciones de entrada/salida, visualizar redes y numerosas funcionalidades más. Todos los datos y parámetros se encuentran disponibles tanto para lectura como para escritura. Algunas de las tareas que se pueden realizar con está interfaz son:

- **caffe.Net** es la interfaz central para cargar, configurar y ejecutar modelos.

- **caffe.Classifier y caffe.Detector** proporcionan interfaces para tareas de clasificación y detección.
- **caffe.SGDSolver** se trata de la interfaz de resolución.
- **caffe.io** maneja funciones de entrada/salida con preprocesamiento.
- **caffe.draw** visualiza las arquitecturas de red.

### 2.1.3. Capas de una red neuronal

Para crear un modelo de Caffe es necesario definir la arquitectura del mismo utilizando para ello un archivo de definición de buffer de protocolo (prototxt).

#### Capas de datos

Los datos entran en Caffe a través de las capas de datos las cuáles se encuentran en la parte inferior de las redes. Estos datos pueden provenir de bases de datos (LevelDB o LMDB), directamente de la memoria, o, cuando la eficiencia no es crítica, desde archivos en disco en formato HDF5 o formatos de imagen comunes.
Tareas comunos de preprocesamiento de los datos de entrada, tales como escalado o reflejo, están disponibles especificando *TransformationParameters* por algunas de las capas. Los tipos de capas "bias", "scalez çrop"pueden ser útiles para el preprocesamiento de la entrada cuando la opción *TransformationParameters* no está disponible.

Capas:

- **Image Data:** Lee imágenes sin procesar.
- **Database:** Lee los datos de LevelLDB o LMDB.
- **HDF5 Input:** Lee los datos en formato HDF5 permitiendo que estos tengan dimensiones arbitrarias.
- **HDF5 Output:** Escribe datos en formato HDF5
- **Input:** Normalmente utilizada para redes que se están implementando.
- **Window Data:**
- **Memory Data:** Lee archivos directamente desde memoria.
- **Dummy Data:** Utilizado para datos estáticos.

#### Capas de visión

Las capas de visión, generalmente toman imágenes como datos de entradas y generan otras imágenes como salida aunque también pueden tomar datos de otros tipos y dimensiones. Una imagen puede tener un canal ($c = 1$) si se trata de una imagen en escala de grises o 3 canales ($c = 3$) si se trata de una imagen RGB. Pero en este contexto, las características distintivas para el tratamiento

de las imágenes de entrada serán la altura y la anchura de las mismas. La mayoría de las capas de visión trabajan aplicando una operación particular sobre alguna región de la entrada para producir una región correspondiente a la salida.

Capas:

- **Convolution Layer:** Convoluciona la imagen de entrada con un conjunto de filtros.
- **Pooling Layer:** Realiza *pooling* de los datos de entrada utilizando para ello funciones de máximo, media o estocásticas.
- **Spatial Pyramid Pooling (SPP)**
- **Crop**
- **Deconvolution Layer:** Realiza uno convolución transpuesta.
- **Im2Col**

#### Capas recurrentes

Capas:

- **Recurrent**
- **RRNN**
- **Long-Short Term Memory (LSTM)**

#### Capas comunes

Capas:

- **Inner Product:** Capa totalmente conectada
- **Dropout**
- **Embed**

#### Capas de pérdida

Estas capas de pérdida conducen al aprendizaje comparando la salida obtenida con el valor de la entrada asignado así un coste para minimizarla.

Capas:

- **Multinomial Logistic Loss**
- **Infogain Loss**
- **Softmax with Loss**

- **Sum-of-Squares / Euclidean**
- **Hinge / Margin**
- **Sigmoid Cross-Entropy Loss**
- **Accuracy / Top-k layer**
- **Contrastive Loss**

#### Capas de normalización

Capas:

- **Local Response Normalization (LRN):** Normaliza regiones locales de los datos de entrada.
- **Mean Variance Normalization (MVN):** Realiza una normalización de contraste / normalización de instancia.
- **Batch Normalization:** Realiza normalizaciones sobre pequeños lotes de datos de entrada.

#### Capas de activación

En general, estas capas son operados que toman un dato de la salida de la capa anterior y generan datos son las mismas dimensiones.

Capas:

- **ReLU / Rectified-Linear and Leaky-ReLU**
- **PReLU**
- **ELU**
- **Sigmoid**
- **TanH**
- **Absolute Value**
- **Power**

$$f(x) = (shift + scale * x)^{power} \tag{2.1}$$

- **Exp**

$$f(x) = base^{(shift+scale*x)} \tag{2.2}$$

- **Log**

$$f(x) = log(x) \tag{2.3}$$

- **BNLL**

$$f(x) = log(1 + exp(x)) \tag{2.4}$$

- **Threshold**: Realiza la función de paso en el umbral definido por el usuario.
- **Bias**
- **Scale**

**Capas de utilidad**

Capas:

- **Flatten**
- **Reshape**
- **Batch Reindex**
- **Split**
- **Concat**
- **Slicing**
- **Eltwise**
- **Filter / Mask**
- **Parameter**
- **Reduction**
- **Silence**
- **ArgMax**
- **Softmax**
- **Python**

## 2.2. JdeRobot

Se trata de un framework cuyo objetivo es desarrollar aplicaciones en robótica y visión por computadora. También tiene actuación en domótica y en escenarios con sensores, accionadores y software inteligente. Ha sido desarrollado para ayudar en la programación de este software inteligente. Está escrito principalmente utilizando el lenguaje C++ proporcionando un entorno de programación en el que el programa de aplicación está compuesto por una colección de varios componentes asíncronos concurrentes. Estos componentes pueden ejecutarse en

diferentes equipos y están conectados mediante el middleware de comunicaciones ICE. Los componentes pueden estar escritos en C++, Python, Java y todos ellos ellos interactúan a través de interfaces ICE explícitas.

JdeRobot simplifica el acceso a dispositivos hardware desde el programa de control. Obtener mediciones de sensores es tan simple como llamar a una función local y ordenar comandos de motor tan fácil como llamar a otra función local. La plataforma adjunta esas llamadas a invocaciones remotas sobre los componentes conectados al sensor o los dispositivos de accionamiento. También, pueden conectarse a sensores y activadores reales o simulados, tanto a nivel local como remoto utilizando para ello la red. Esas funciones construyen la API para la capa de abstracción del hardware. La aplicación robótica obtiene las lecturas del sensor y ordena los comandos del actuador usando esa API para desplegar su comportamiento. Se han desarrollado varios drivers para soportar diferentes sensores, activadores y simuladores. Los robots y sensores actualmente soportados son:

- **Sensores RGBD:** Kinect and Kinect2 de Microsoft, Asus Xtion
- **Robots con ruedas:** Kobuki (TurtleBot) de Yujin Robot y Pioneer de MobileRobotics Inc.
- **ArDrone quadrotor de Parrot**
- **Escáneres laser:** LMS de SICK, URG de Hokuyo y RPLidar
- **Simulador Gazebo**
- **Cámaras Firewire, cámaras USB, archivos de vídeo (mpeg, avi), cámaras IP (como Axis)**

JdeRobot incluye varias herramientas de programación de robots y bibliotecas. En primer lugar, teleespectadores y teleoperadores para varios robots y sus sensores y motores. En segundo lugar, un componente de calibración de cámara y una herramienta de tunning para filtros de color. En tercer lugar, una herramienta llamada VisualHFSM para la programación del comportamiento del robot utilizando la jerarquía Finite State Machines. Además, también proporciona una biblioteca para desarrollar controladores difusos y otra para la geometría proyectiva y el procesamiento de la visión por computadora.

Cada componente puede tener su propia interfaz gráfica de usuario o ninguna en absoluto. Actualmente, las bibliotecas GTK y Qt son compatibles, incluyéndose varios ejemplos de OpenGL para gráficos 3D con ambas bibliotecas.

JdeRobot es un software de código abierto con licencia como GPL y LGPL. También utiliza software de terceros como el simulador Gazebo, ROS, OpenGL, GTK, Qt, Player, Stage, GSL, OpenCV, PCL, Eigen u Ogre.

## 2.3. Base de datos LMDB

Lightning Memory-Mapped Database (LMDB) es una biblioteca de software que proporciona una base de datos de clave-valor de alto rendimiento. LMDB

almacena los pares de datos de forma arbitraria utilizando arrays de bytes.

### 2.3.1. LMDB con Python

Es posible trabajar fácilmente con bases de datos LMDB utilizando Python. En primer lugar, es necesario importar la librería de LMDB.

```
import lmdb
```

Posteriormente, se pueden utilizar los siguientes comandos.

```
lmdb_env = lmdb.open('Ruta donde se encuentra la base de datos')
lmdb_txn = lmdb_env.begin()
lmdb_cursor = lmdb_txn.cursor()
datum = caffe.proto.caffe_pb2.Datum()
```

De esta manera tendremos a disposición los datos del archivo LMDB en formato Datum. Datum es una clase que almacena datos y opcionalmente etiquetas. Estos datos que guarda se representan con un array de 3 dimensiones: altura, anchura y canal. Cada valor del array corresponde con el nivel de intensidad del píxel al que hace referencia.

# Capítulo 3

# Clasificación

## 3.1. Red Neuronal

### 3.1.1. Estructura de la red

En esta sección se explicará la estructura de la red que utilizaremos para la clasificación de los dígitos manuscritos que nos facilita la base de datos MNIST.

En primer lugar, se especificará el nombre de la red, en este caso "LeNet"

```
name: "LeNet"
```

Posteriormente, es necesario leer los datos de la base de datos LMDB. Para ello, se define la capa de datos:

```
layer {
  name: "mnist"
  type: "Data"
  transform_param {
    scale: 0.00390625
  }
  data_param {
    source: "mnist_train_lmdb"
    backend: LMDB
    batch_size: 64
  }
  top: "data"
  top: "label"
}
```

La capa de convolución tomará los datos de la capa de entrenamiento y los transformará utilizando un núcleo de convolución de 5X5, produciendo 20 salidas. Para inicializar los pesos se utilizará el algoritmo Xavier, el cuál determina automáticamente la escala de inicialización basándose en el número de entradas

y salidas de cada neurona. Adicionalmente, el sesgo se inicializará como una constante, siendo esta por defecto 0.

```
layer {
  name: "pool1"
  type: "Pooling"
  pooling_param {
    kernel_size: 2
    stride: 2
    pool: MAX    }
  bottom: çonv1"
  top: "pool1"
}
```

```
layer {
  name: ïp1"
  type: ÏnnerProduct"
  param { lr_mult: 1 }
  param { lr_mult: 2 }
  inner_product_param {
    num_output: 500
    weight_filler {
      type: "xavier"
    }
    bias_filler {
      type: çonstant"
    }
  }
bottom: "pool2"
top: ïp1"
}
```

```
layer {
  name: ïp1”
  type: ÏnnerProduct”
  param { lr_mult: 1 }
  param { lr_mult: 2 }
  inner_product_param {
    num_output: 500
    weight_filler {
      type: ”xavier”
    }
    bias_filler {
      type: çonstant”
    }
  }
  bottom: ”pool2”
  top: ïp1”
}
```

```
layer {
  name: relu1”
  type: ReLU”
  bottom: ïp1”
  top: ïp1”
}
```

```
layer {
  name: ”loss”
  type: ”SoftmaxWithLoss”
  bottom: ïp2”
  bottom: ”label”
}
```

En la siguiente imagen se observa la estructura de la red explicada anteriormente:

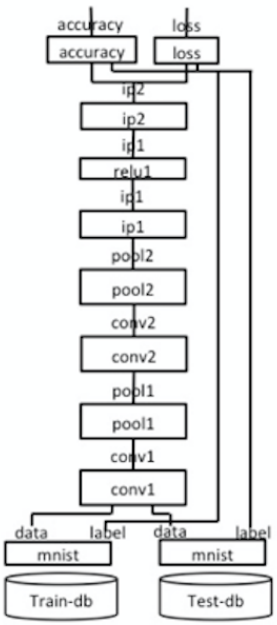


Figura 3.1: Estructura red neuronal

### 3.1.2. Definición del solucionador

El solucionador es el responsable de la optimización del modelo. Se define en un archivo con extensión *.prototxt*, en este caso, *lenet_solver.prototxt*.
Este solucionador calcula la precisión del modelo usando el conjunto de validación cada 100 iteraciones. El proceso de optimización tendrá una duración de máximo 10000 iteraciones y tomará una instantánea del modelo entrenado cada 500 iteraciones.
En esta configuración, se empezará con una tasa de aprendizaje de 0.01 (*base_lr = 0.01*), la cual irá cayendo con un factor de 10000 (*gamma = 0.0001*).
El solucionador *lenet_solver.prototxt* tendrá el siguiente aspecto. Se puede observar un comentario antes de cada sentencia indicando su funcionalidad:

```
# The train/test net protocol buffer definition
net: .examples/mnist/lenet_train_test.prototxt"
# test_iter specifies how many forward passes the test should carry out.
# In the case of MNIST, we have test batch size 100 and 100 test iterations,
# covering the full 10,000 testing images.
test_iter: 100
# Carry out testing every 500 training iterations.
test_interval: 500
# The base learning rate, momentum and the weight decay of the network.
base_lr: 0.01
momentum: 0.9
weight_decay: 0.0005
# The learning rate policy
lr_policy: ïnv"
gamma: 0.0001
power: 0.75
# Display every 100 iterations
display: 100
# The maximum number of iterations
max_iter: 10000
# snapshot intermediate results
snapshot: 5000
snapshot_prefix: .examples/mnist/lenet"
# solver mode: CPU or GPU
solver_mode: GPU
```

### 3.1.3. Entrenamiento de la red

Tras la definición de la red y el solucionador, para entrenar el modelo hay que ejecutar, en la ruta donde se encuentre el directorio de caffe, el siguiente comando.

```
cd $CAFFE_ROOT
./examples/mnist/train_lenet.sh
```

El script *train_lenet.sh* hará una llamada al solucionador que se quiera ejecutar.

```
#!/usr/bin/env sh


./build/tools/caffe train –solver=examples/mnist/lenet_solver.prototxt
```

Una vez entrenada la red, se generarán 2 ficheros:

- lenet_iter_10000.caffemodel: Es un binario que contiene el estado actual de los pesos para cada capa de la red.
- lenet_iter_10000.solverstate: Es un binario que contiene la información necesaria para continuar el entrenamiento del modelo desde donde se detuvo por última vez.

## 3.2. Base de datos MNIST

La base de datos MNIST (Modified National Institute of Standards and Technology database) es una gran base de datos de dígitos escritos a mano que se utiliza comúnmente para el entrenamiento de sistemas de procesamiento de imágenes.

Esta base de datos está compuesta por 60000 imágenes destinadas al entrenamiento y 10000 imágenes de prueba. Las imágenes están representadas en escala de grises con 256 niveles de intensidad y con un tamaño normalizado de 28x28 píxeles. Cada imagen está formada por un dígito escrito a mano comprendido entre el 0 y el 9, centrado y con el fondo negro, es decir, con un nivel de intensidad igual a 0.

### 3.2.1. Sección de datos

Para entender mejor con que datos se va a trabajar, es necesario visualizar algún ejemplo de la base de datos MNIST. Con este fin, utilizaremos el formato PGM (Portable Graymap Format) para representar las imágenes. Para conseguir una imagen .pgm hay que crear un documento .txt que tenga el siguiente formato:

- En la primera línea escribir el código **P2** para identificar que lo que se quiere conseguir es una imagen con extensión pgm.
- En la segunda línea el ancho y el alto de la imagen, en nuestro caso **28 28**.
- En la tercera línea el número de grises entre el blanco y el negro, en nuestro caso **255**.
- Por último, escribir la intensidad de cada uno de los píxeles de la imagen perfectamente alineados.

La siguiente imagen muestra un ejemplo de una imagen en formato PGM.

```
P2
28 28    # Width and Height.
255       # Numbers of grey between black and white.

0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   3  18  18  18 126 136 175  26 166 255 247 127   0   0   0   0
0   0   0   0   0   0   0   0  30  36  94 154 170 253 253 253 253 253 225 172 253 242 195  64   0   0   0   0
0   0   0   0   0   0   0  49 238 253 253 253 253 253 253 253 253 251  93  82  82  56  39   0   0   0   0   0
0   0   0   0   0   0   0  18 219 253 253 253 253 253 198 182 247 241   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0  80 156 107 253 253 205  11   0  43 154   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0  14   1 154 253  90   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0 139 253 190   2   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0  11 190 253  70   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0  35 241 225 160 108   1   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0  81 240 253 253 119  25   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0  45 186 253 253 150  27   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0  16  93 252 253 187   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 249 253 249  64   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0  46 130 183 253 253 207   2   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0  39 148 229 253 253 253 250 182   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0  24 114 221 253 253 253 253 201  78   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0  23  66 213 253 253 253 253 198  81   2   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0  18 171 219 253 253 253 253 195  80   9   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0  55 172 226 253 253 253 253 244 133  11   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0 136 253 253 253 212 135 132  16   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
```

Figura 3.2: Fichero PGM

Para crear este archivo se utilizará lo indicado en la sección 2.3, obteniendo así un array de 3 dimensiones con los niveles de intensidad de cada uno de los píxeles de cada imagen de la base de datos.

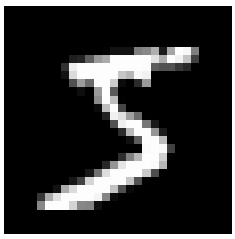

Figura 3.3: Imagen base de datos MNIST

### 3.2.2. Sección de etiquetas

Zero: 5923
One: 6742
Two: 5958
Three: 6131
Four: 5842
Five: 5421
Six: 5918
Seven: 6265
Eight: 5851
Nine: 5949
Total images: 60000

# Bibliografía

Edgar Nelson Sánchez Camperos y Alma Yolanda Alanís García: Redes Neuronales: Conceptos fundamentales y aplicaciones a control automático"