Predicción de la Velocidad del Tráfico Basada en Redes Neuronales Convolucionales

Jesús Mena-Oreja y Javier Gozalvez[1]

Resumen—**La congestión del tráfico rodado tiene un importante impacto socioeconómico y medioambiental. Una gestión dinámica y precisa del tráfico permitiría un uso más eficiente de la infraestructura vial y prevenir situaciones de congestión. Esta gestión sería posible si se dispusiera de una predicción precisa del estado del tráfico, para lo cual este artículo propone una nueva técnica de predicción de tráfico basada en redes neuronales convolucionales. A diferencia de las soluciones existentes, estas redes permiten explotar no sólo la evolución temporal del tráfico, sino también su evolución espacial. La red propuesta ha sido entrenada utilizando datos reales de tráfico, y su rendimiento ha sido comparado con soluciones actuales, en concreto con el de una red neuronal que solo procesa temporalmente los datos de tráfico. Los resultados demuestran que el procesado espaciotemporal de los datos de tráfico mediante la red neuronal convolucional propuesta mejora la predicción del estado del tráfico un 22% en términos de error absoluto y un 35% en términos de error relativo.**

Palabras clave—**Predicción de tráfico; Redes Neuronales; Redes Neuronales Convolucionales; Aprendizaje Automático; Aprendizaje Profundo; Sistemas de Transporte Inteligentes; ITS**

I. INTRODUCCIÓN

La congestión del tráfico y los atascos tienen un impacto significativo en la economía y el confort de los conductores. Los conductores pasan millones de horas en atascos cada año y eso conlleva enormes pérdidas económicas así como otros problemas asociados como la contaminación o accidentes de tráfico. La Comisión Europea cifra el coste asociado a la congestión en la Unión Europea en 100 billones de euros (un 1% del PIB) [1]. Una gestión del tráfico y la infraestructura vial dinámica, eficiente y sostenible permitiría paliar estos efectos. Para ello es necesario conocer las características del tráfico e incluso poder predecirlas para poder anticipar problemas. Disponer de información sobre el estado actual o futuro del tráfico permite a los gestores de tráfico tomar decisiones que ayudan a mitigar los problemas asociados a la congestión, y a los conductores planificar sus viajes buscando la ruta más eficiente y segura. Conocer el estado del tráfico, presente y futuro, es en parte posible gracias a los diferentes tipos de sensores (por ejemplo, espiras, cámaras u otro tipo de aforadores) desplegados a lo largo de las carreteras y al creciente auge de los datos FCD (*Floating Car Data*) proporcionados por los vehículos conectados.

[1]Laboratorio UWICORE, Departamento de Ingeniería de Comunicaciones, Universidad Miguel Hernández de Elche, e-mail: {jmena, j.gozalvez}@umh.es

La capacidad de gestionar de forma dinámica y eficiente el tráfico depende en gran medida de la capacidad de poder predecir su estado futuro a través del análisis de su evolución en instantes anteriores. Varios tipos de técnicas han sido propuestas hasta la fecha para predecir el estado del tráfico, siendo las redes neuronales las más comunes. Sin embargo, las técnicas propuestas hasta la fecha se centran en el estudio de la evolución temporal del tráfico y no en su evolución espacial, a pesar de que el tráfico se caracteriza por su correlación espaciotemporal debido a su comportamiento como onda de choque que se propaga a lo largo de la carretera y a través del tiempo [2].

En este contexto, este artículo presenta una nueva técnica de predicción de tráfico basada en redes neuronales convolucionales (CNN, *Convolutional Neural Networks*) [3]. La técnica propuesta predice, al conocimiento de los autores por primera vez, el estado del tráfico teniendo en cuenta su evolución temporal y espacial a lo largo de la carretera. El estudio demuestra que la técnica propuesta mejora la precisión de la predicción del estado del tráfico con respecto a redes neuronales que sólo tienen en cuenta la evolución temporal del tráfico. Es importante destacar que ambas redes han sido entrenadas con datos reales de tráfico proporcionados por el Departamento de Transporte de California a través de su plataforma PeMS (*Performance Measurement System*) [4]. PeMS es la base de datos de sensores de tráfico más extensa y actualizada que proporciona diversas métricas como intensidad, velocidad y densidad del tráfico.

El contenido del presente artículo se organiza de la siguiente manera. La sección II analiza los principales estudios científicos que tratan el problema de predicción de tráfico. A continuación, la sección III presenta el modelo de red neuronal convolucional propuesto en este trabajo. La sección IV describe el conjunto de datos utilizado para entrenar el modelo propuesto. El rendimiento de dicho modelo es analizado en la sección V. Finalmente, la sección VI presenta las conclusiones de este estudio y posibles líneas futuras de este trabajo.

II. TRABAJOS RELACIONADOS

El problema de predicción de tráfico ha sido abordado en la literatura con diferentes enfoques y técnicas que son recopilados en los trabajos [5] y [6]. La mayoría de los enfoques propuestos hasta la fecha usa modelos que utilizan como variables de entrada una lista de las variables de tráfico en varios instantes de tiempo pasados. Los modelos más utilizados para predicción de tráfico son los basados en redes neuronales, habiéndose propuesto en la literatura diferentes tipos de

arquitecturas. Por ejemplo, [7] propone el uso de un perceptrón multicapa (MLP, *Multilayer Perceptron*) para predecir valores futuros de la intensidad del tráfico usando como entrada valores pasados de la intensidad. El trabajo expuesto en [8] también recurre a un MLP, aunque en este caso para predecir la velocidad de los vehículos en diferentes tramos de una carretera. Para predecir la velocidad en un tramo, los autores emplean valores de velocidad medidos en instantes anteriores en dicho tramo y en los tramos anterior y posterior. Este trabajo es uno de los primeros estudios que incorpora información espacial y temporal en la entrada de la red neuronal con el fin de mejorar la precisión de la predicción. Otras propuestas de MLP consideran como entrada a la red neuronal variables obtenidas de un preprocesado de la evolución temporal de las variables de tráfico en lugar de los valores concretos de las variables [9]. Otro tipo de redes neuronales utilizadas para predicción de tráfico son las redes neuronales recurrentes y las redes recurrentes basadas en LSTM (*Long Short-Term Memory*). El estudio presentado en [10] compara la precisión obtenida por diferentes redes neuronales recurrentes (en concreto, las redes de Elman, Jordan y la red neuronal de espacio de estado) en la predicción de los tiempos de recorrido. Los trabajos presentados en [11] y [12] utilizan redes LSTM para predecir la velocidad y la intensidad del tráfico, respectivamente. Las redes de tipo recurrente sólo necesitan como entrada el último valor de las variables de tráfico en cada instante. Existen otras propuestas que conviene también destacar como son los autoencoders [13] o redes con arquitecturas híbridas que utilizan el perceptrón multicapa junto a otras técnicas [14][15].

Otro tipo de técnicas muy utilizadas en predicción de tráfico son los modelos paramétricos. Dentro de esta categoría se pueden encontrar técnicas como los modelos autoregresivos o el filtro de Kalman. Los modelos autoregresivos modelan la variable de salida como una función de sus valores pasados. Los modelos autoregresivos más utilizados en predicción de tráfico son el ARIMA (*AutoRegressive Integrated Moving Average*) y el SARIMA (*Seasonal ARIMA*). Estos modelos han sido utilizados en [16] y [17] para modelar la intensidad del tráfico. El filtro de Kalman ha sido utilizado para predecir los tiempos de recorrido modelando las variables de tráfico como un sistema dinámico en el espacio de estados [18] o fusionando la información proveniente de diferentes sensores [19].

Otras técnicas propuestas en la literatura para predecir variables de tráfico son el algoritmo de los k vecinos más cercanos [20][21], las redes bayesianas [22], y la regresión con vectores de soporte [23].

Tal y como se puede observar en [5] y [6], los trabajos que mejores resultados obtienen al predecir valores futuros de variables de tráfico son aquellos basados en redes neuronales. Además, las redes neuronales son el tipo de técnica más utilizado para predicción de tráfico y son muy versátiles y robustas ante el ruido. Por todo esto, este artículo se centra en el estudio de arquitecturas de redes neuronales para predicción de tráfico.

Ninguno de los estudios mencionados anteriormente explota la evolución espaciotemporal del tráfico (sólo la temporal) para predecir su estado futuro. Si bien [8] utiliza la información de los tramos de carretera adyacentes al tramo de estudio, el modelo utilizado no está específicamente diseñado para procesar patrones multidimensionales, como los que puede presentar el tráfico.

III. MODELO DE RED NEURONAL CONVOLUCIONAL

Este artículo propone una técnica basada en redes neuronales convolucionales que es capaz de procesar patrones espaciotemporales con el fin de mejorar la precisión de la predicción del tráfico. Las redes neuronales convolucionales son utilizadas frecuentemente en visión por computador, y han demostrado que son superiores a otras técnicas en la detección de características de una imagen que pueden determinar los objetos que se encuentran en ésta, sin importar la localización o el contexto en que se encuentren dichos objetos. Por ejemplo, si en una imagen hay un gato, la red convolucional detectará el gato si se encuentra en una de las cuatro esquinas, en el centro o en cualquier región de la imagen. Si se trata la evolución del tráfico como una imagen (la dimensión espacial en una dimensión de la imagen y la dimensión temporal en la otra), y los eventos de tráfico como una marca en la imagen, las redes neuronales convolucionales representan una interesante opción para intentar detectar esos eventos, sin importar dónde y cuándo ocurran. Esta característica de las redes convolucionales podría mejorar la precisión de la predicción del tráfico.

A. Redes neuronales convolucionales

Las redes neuronales se basan en el modelo de neurona artificial, el cual está basado en el modelo de las neuronas biológicas. Una neurona artificial recibe una serie de entradas. La neurona realiza una suma ponderada de las distintas entradas, añadiendo además un valor umbral a dicha suma. La salida de la neurona es el resultado de aplicar una función llamada función de activación al resultado de sumar el umbral a la suma ponderada de las entradas. El cálculo de la salida de una neurona se realiza según la Ecuación (1):

$$y = \mathcal{F}\left(\sum_{i=1}^{N} w_i x_i + b\right) \tag{1}$$

Donde x_i es cada una de las entradas a la neurona, w_i es el peso que pondera la entrada x_i, N es el número de entradas a la neurona, b es el umbral de la neurona, $\mathcal{F}$ es la función de activación de la neurona, que generalmente es una función no lineal, e y es la salida de la neurona.

Las redes neuronales consisten, de manera general, en una serie de capas de neuronas apiladas unas encima de otras. Una capa está formada por un conjunto de neuronas que procesan las entradas de la capa, que es la salida de las neuronas de la capa anterior, excepto en el caso de la primera capa de la red o capa de entrada, cuya entrada es la entrada de la red. La última capa de la red se llama capa de salida, y las capas que se encuentran entre la capa de entrada y la capa de salida se conocen como capas ocultas.

Las redes neuronales son consideradas aproximadores universales porque se pueden ajustar sus parámetros para aproximar cualquier función. El proceso mediante

el cual se ajustan los parámetros de una red neuronal se llama entrenamiento de la red, y los parámetros entrenables de la red son los pesos y umbrales de las neuronas de todas las capas que forman la red neuronal. Este proceso de entrenamiento consiste en minimizar una función de coste, generalmente una medida del error cometido en la salida de la red, ajustando los parámetros entrenables de la red. Este problema de minimización es muy difícil de resolver analíticamente, por lo que se recurre a métodos iterativos como el método del descenso del gradiente o métodos derivados de éste. Los métodos de descenso del gradiente consisten en moverse por el espacio de los parámetros entrenables en la dirección negativa del gradiente de la función de coste respecto de los parámetros entrenables:

$$\theta \leftarrow \theta - \alpha \frac{\partial \mathcal{L}}{\partial \theta} \tag{2}$$

Donde θ es un vector que incluye todos los parámetros entrenables de la red (pesos y umbrales), $\mathcal{L}$ es la función de coste, y α es la tasa de aprendizaje. La tasa de aprendizaje es un parámetro que controla el ritmo al que se aprende en el proceso de entrenamiento. Una tasa de aprendizaje alta hará que el entrenamiento converja más rápido, pero el ajuste de los parámetros entrenables será menos preciso y dará lugar a tasas de error más altas. En cambio, cuando la tasa de aprendizaje es baja se consiguen errores más bajos, pero el entrenamiento es más lento. La estrategia general para la tasa de aprendizaje consiste en reducir su valor a lo largo del entrenamiento para tener una convergencia mayor al principio y errores más bajos al final.

Cada neurona de una red neuronal se encarga de detectar una característica o patrón de la entrada de la red neuronal. Cuantas más capas ocultas tiene una red neuronal más complejas son las características que las neuronas son capaces de detectar. La salida de la red neuronal es una función de las características detectadas por las capas ocultas. Las características que detectarán las neuronas de la red no están determinadas antes del entrenamiento, y es la propia red la que "decide" durante el entrenamiento qué clase de características buscar.

Las redes neuronales convolucionales son un caso particular de redes neuronales que contienen una o más capas convolucionales. Estas capas convolucionales se caracterizan por compartir sus parámetros entrenables a lo largo de sus entradas. Esto se consigue mediante el uso de filtros de convolución que se aplican sobre las entradas de la capa. En las capas convolucionales, en lugar de aprender un peso por cada entrada, se aprenden los valores de un filtro de convolución que es común a toda la entrada y que se desliza por toda la entrada para calcular la salida. Además de los filtros de convolución, las capas convolucionales comparten también un umbral por cada filtro de convolución. Cada vez que se desliza el filtro de convolución, éste se aplica sobre una subregión de la entrada del mismo tamaño que el filtro. Esto es equivalente a tener una red neuronal pequeña que se aplica sobre las distintas subregiones de la imagen y calcula una salida para cada subregión. Si el filtro de convolución es del mismo tamaño que la entrada a la capa el resultado no difiere del de una capa normal. En el caso en el que la entrada a la capa sea una imagen (con una altura, una anchura y un número de canales determinado) los filtros de convolución serán bidimensionales y se deslizarán a lo alto y ancho de la imagen. La salida de una capa convolucional es otra imagen con tantos canales como filtros de convolución aplique la capa.

En general, la arquitectura de una red neuronal convolucional que trabaja con imágenes consiste en una serie de capas convolucionales apiladas, de manera que cada capa aplique los filtros de convolución aprendidos sobre la salida de la capa anterior, y en el caso de la capa de entrada los filtros se aplican sobre la imagen de entrada. Tras las capas convolucionales suele haber una serie de capas no convolucionales, también llamadas densas o totalmente conectadas (FC, *Fully Connected*). Estas capas densas son iguales que las capas ocultas de una red neuronal normal. La Figura 1 muestra la arquitectura general de una red neuronal convolucional que trabaja con imágenes.

Cada filtro de convolución aprendido por una red neuronal convolucional está especializado en detectar una característica. Las características detectadas por los filtros de las redes convolucionales siguen la misma línea que las de una red no convolucional, no están determinadas antes del entrenamiento y en general son características poco intuitivas. Aun así, en las primeras capas de las redes convolucionales utilizadas en visión por computador es común encontrar filtros especializados en detectar ciertos patrones geométricos o relacionados con el color, por ejemplo líneas horizontales. Las características detectadas por una red neuronal convolucional pueden ser detectadas en cualquier lugar de la imagen de entrada, pues los parámetros entrenables son compartidos espacialmente por toda la imagen. Por esta cualidad se suele decir que las redes neuronales convolucionales poseen invarianza traslacional. La salida de las capas convolucionales es una imagen cuyos canales son mapas de características. Las capas densas de una red convolucional trabajan con los mapas de características de la última capa convolucional para determinar el objeto que se encuentra en la imagen o dar una salida en función de sus características.

B. Formato de la entrada de la red

Los datos de tráfico están distribuidos espacialmente y temporalmente, y están formados por diversas variables de tráfico que caracterizan el estado del tráfico en cada tramo y en cada instante. En este trabajo las variables de tráfico consideradas son las tres variables fundamentales del tráfico, la intensidad del tráfico (el número de vehículos que circulan por la carretera por unidad de tiempo), la velocidad media de los vehículos, y la densidad del tráfico (el número de vehículos que se encuentran en la carretera por unidad de longitud). Una posibilidad sería utilizar las variables de tráfico que proporciona cada sensor como un canal de la señal de entrada a la red y aplicar convoluciones unidimensionales a la evolución temporal de cada variable de cada sensor por separado, tal y como se hace en [24]. El inconveniente de este procedimiento es que no se explotarían las relaciones espaciotemporales existentes en el tráfico. Para poder aprovechar estas relaciones, este trabajo propone utilizar como canales de la señal de entrada cada una de las variables de tráfico.

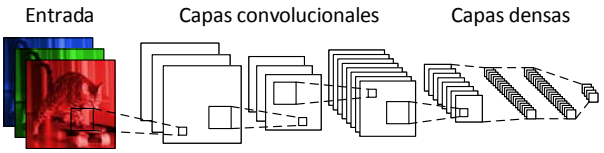


Fig. 1: Arquitectura general de una red convolucional.

Cada canal está formado por una matriz cuyas filas y columnas incluyen, respectivamente, la información temporal y espacial de la variable. De esta manera, se forma una imagen del tráfico en la que cada píxel corresponde a los valores de las variables en cuestión en una ubicación concreta (la del sensor correspondiente) y en un momento dado. El alto de la imagen se corresponde con el número de sensores seleccionados y el ancho con la ventana temporal utilizada para la predicción, es decir, el número de instantes anteriores tenidos en cuenta para la predicción. La red dispone pues de tres canales de entrada que se corresponden con la evolución de las tres variables de tráfico objeto de estudio (intensidad, velocidad media y densidad) a lo largo de todo el tramo de estudio y a lo largo de la ventana temporal. Representar la evolución espaciotemporal del tráfico como una imagen no es algo nuevo, pues en el PeMS se pueden obtener gráficas de la evolución de las variables de tráfico en forma de imagen. La novedad que se propone en el presente artículo es aprovechar esta representación de los datos de tráfico para procesarlos con redes neuronales convolucionales usando filtros de convolución bidimensionales de manera que se aprovechen los patrones espaciotemporales presentes en la evolución del tráfico.

C. Propuesta de red neuronal convolucional para predicción de tráfico

La arquitectura de la red propuesta consiste, igual que en el resto de redes neuronales convolucionales, en una serie de capas apiladas que procesan la información que reciben como entrada. La primera capa de la red es la capa de entrada, que únicamente normaliza la entrada de la red de acuerdo con la Ecuación (3) para que los valores de la entrada se encuentren distribuidos con media cero y varianza unidad:

$$\bar{x} = \frac{x - \mu}{\sigma} \tag{3}$$

Donde x es el vector de entrada sin normalizar, $\bar{x}$ es el vector de entrada normalizado, μ es la media de los vectores de entrada del conjunto de entrenamiento, y σ es la desviación típica de los vectores de entrada del conjunto de entrenamiento. La división de la Ecuación (3) es una división elemento a elemento. Con la normalización de la entrada los valores de la entrada no toman valores ni muy altos ni muy bajos, con lo que se consigue que el entrenamiento de la red sea más estable.

Tras la capa de entrada se encuentran una serie de capas convolucionales apiladas que se encargan de realizar un procesado espaciotemporal de los datos de tráfico para detectar características del tráfico. Cada capa procesa la salida de la capa anterior, mientras la primera capa convolucional procesa los datos de tráfico normalizados. Las capas convolucionales han sido diseñadas siguiendo los principios de diseño presentados en [25]. En concreto se han usado sólo filtros de convolución de tamaño 3x3 apilados unos encima de otros manteniendo la función de activación tras cada convolución. Tal y como se muestra en [25] dos filtros de tamaño 3x3, con 18 parámetros entrenables (9 de cada filtro) consiguen un procesado de su entrada equivalente al de un filtro de tamaño 5x5, con 25 parámetros entrenables; tres filtros de tamaño 3x3 apilados consiguen un procesado equivalente al de un filtro de tamaño 7x7, y así sucesivamente. [25] también demuestra que con esta estrategia se consigue el mismo rendimiento o mejor con un menor número de parámetros entrenables que con filtros de mayor tamaño, consiguiendo así una arquitectura de red neuronal más eficiente en términos de memoria y de cómputo. Todas las capas convolucionales usan como función de activación la función ReLU (*Rectified Linear Unit*) [26], mostrada en la Ecuación (4):

$$\mathrm{ReLU}(x) = \max(0, x) \tag{4}$$

Donde x es la variable de entrada de la función ReLU. La función ReLU ha demostrado su superioridad con respecto a otras funciones de activación como la sigmoidea o la tangente hiperbólica en multitud de campos, entre ellos el de la visión por computador. En este trabajo también se han probado las funciones de activación sigmoidea y tangente hiperbólica y los resultados son inferiores a los obtenidos usando la función ReLU.

Tras las primeras pruebas se detectó que la predicción no sólo no mejoraba con el uso de capas convolucionales, sino que al aumentar el número de capas la precisión de la predicción empeoraba. Tal y como se indica en [27], cuanto más profunda es una red neuronal, es decir, cuantas más capas tiene, más complejas son las características que es capaz de detectar y por tanto mayor debe ser su precisión. Por tanto, la razón por la que el aumento del número de capas convolucionales empeora la predicción puede ser debido a dos razones distintas, que el uso de capas convolucionales no ayude en la predicción de tráfico, o que su uso complica el entrenamiento de la red hasta tal punto que no se consigue encontrar una red que resuelva el problema de predicción. En [28] se analiza el problema de la dificultad del entrenamiento de redes neuronales convolucionales cuando la profundidad aumenta mucho y proponen el uso de conexiones residuales. Estas conexiones residuales consisten en sumar a la salida de un grupo de capas convolucionales la entrada de la primera capa de dicho grupo. El esquema utilizado en este trabajo se muestra en la Figura 2. El objetivo de esta técnica es que si el grupo de capas no aporta ninguna mejora a la red, mediante la conexión residual la red neuronal tiene la posibilidad de aprender la función identidad, de manera que aunque las capas no aportan ninguna mejora, tampoco empeoran los resultados de la red. Esto permite aumentar la profundidad de la red neuronal sin perjudicar el entrenamiento de ésta. El uso de conexiones residuales en la red neuronal convolucional para predicción de tráfico soluciona el problema del empeoramiento de la precisión al añadir capas convolucionales y se consiguen resultados mejores que sin usar capas convolucionales. Además de usar conexiones residuales, la red aquí

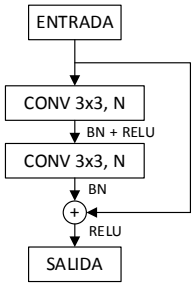


Fig. 2: Esquema de un bloque residual, con *N* el número de filtros de convolución de cada capa.

propuesta también hace uso de la técnica de normalización por lotes (BN, *Batch Normalization*) [29], que ayuda a facilitar el entrenamiento de la red al normalizar la salida de cada filtro de convolución. Esta normalización busca que la red trabaje con valores centrados en el origen y con varianza unidad, en la misma línea que la normalización de la entrada de la red pero extendiéndolo al resto de capas convolucionales de la red. Además, las pruebas realizadas han mostrado que la precisión de la predicción del tráfico mejora con el uso de BN. Finalmente el diseño de las capas convolucionales de la red propuesta queda definido por el uso de 18 capas convolucionales organizadas en bloques residuales como el representado en la Figura 2, que constan de dos capas convolucionales cada uno, con normalización por lotes y una conexión residual, y que son apilados unos encima de otros, de manera que un bloque residual procesa la salida del anterior. Todas las convoluciones se realizan con *same padding*, es decir, aumentando el tamaño de la entrada de cada filtro de convolución con ceros para mantener constante el tamaño de los vectores de entrada a cada capa y no perder información tras cada convolución. No se ha usado ninguna capa de *pooling*, capas que reducen la dimensión (alto y ancho) de su entrada, en ningún lugar de la red, pues se ha observado que empeora los resultados.

Después de las capas convolucionales se encuentran una serie de capas densas, la primera de las cuales toma como entrada la salida del último bloque residual. La finalidad de estas capas es procesar las características del tráfico detectadas por las capas convolucionales para hacer regresión y ajustar una función con la que finalmente obtener la predicción como salida de la red. Todas las capas completamente conectadas usan como función de activación la función ReLU a excepción de la capa de salida, cuya función de activación es la función identidad. De esta forma la salida de cada neurona de la capa de salida es una transformación afín de las activaciones de la capa anterior, como es habitual en las redes neuronales usadas en regresión.

La red neuronal propuesta consta de nueve bloques residuales, tres con capas convolucionales de 32 filtros de convolución, tres con capas de 64 filtros, y tres con capas de 96 filtros. Además, tras las capas convolucionales, la red propuesta consta de tres capas densas, una de 2048 neuronas, otra de 1024 neuronas y la capa de salida con tantas neuronas como variables se deseen predecir. Para evitar el sobreajuste, es decir, que la red aprenda el conjunto de entrenamiento sin generalizar a datos no vistos durante el entrenamiento, se ha usado *dropout* [30], una técnica que consiste en poner a cero la salida de un porcentaje de neuronas, elegidas aleatoriamente, en cada iteración del entrenamiento. Con *dropout* se consigue que las neuronas de la red no cuenten con que otras neuronas vayan a tener una salida distinta de cero, con lo que se evita la coadaptación de neuronas y se asegura una representación redundante de la información de la capa en la que se aplica *dropout*. En la red propuesta se ha usado *dropout* entre la primera y la segunda capa densa con una probabilidad de mantener las activaciones del 60%. No se ha usado *dropout* entre la segunda capa densa y la capa de salida porque durante el entrenamiento la capa de salida no aprendería correctamente la transformación afín de sus entradas. La Figura 3 muestra un esquema de la red propuesta.

D. Entrenamiento de la red

Para entrenar los parámetros de la red se ha recurrido al algoritmo de retropropagación y al método del descenso del gradiente estocástico, entrenando la red con un lote de instancias de entrenamiento en cada iteración en lugar de entrenar con todo el conjunto de entrenamiento, asegurando que el modelo cabe en memoria. El objetivo del entrenamiento es minimizar una función de coste, que en este caso, como se trata de un problema de regresión, es el cuadrado de la norma L2 del error cometido por la red. La expresión de esta función de coste es la de la Ecuación (5):

$$\mathcal{L} = \frac{1}{2}\|\hat{y} - y\|_2^2 \tag{5}$$

Donde $\hat{y}$ es la salida de la red, y es la salida deseada, y $\|\cdot\|_2$ es la norma L2. Minimizar esta función de coste es equivalente a minimizar el error cuadrático cometido por la red.

Para ayudar al entrenamiento y conseguir valores más bajos de la función de coste se ha usado decaimiento exponencial de la tasa de aprendizaje, la cual toma un valor inicial de 10^{-4}, y cada 1000 iteraciones del algoritmo de retropropagación el valor de la tasa de aprendizaje se multiplica por 0.1. Con el decaimiento de la tasa de aprendizaje se consigue una convergencia del algoritmo de retropropagación más rápida al principio del entrenamiento, mientras que al final el ajuste de los parámetros entrenables es más lento pero más preciso al usar una tasa de aprendizaje menor. También se ha limitado la norma del gradiente de la función de coste respecto a los parámetros entrenables a un máximo de 40 para evitar que el entrenamiento oscile sin control al principio. Además, el conjunto de entrenamiento es permutado cambiando el orden de los ejemplos de entrenamiento tras cada iteración con todo el conjunto de entrenamiento entero. Así se evita que los lotes de entrenamiento sean iguales en diferentes iteraciones, de manera que la red no aprenda correlaciones inexistentes entre distintos ejemplos de entrenamiento. El modelo de red neuronal convolucional y el algoritmo de entrenamiento han sido implementados usando el framework TensorFlow [31].

IV. DESCRIPCIÓN DEL CONJUNTO DE DATOS

Las redes neuronales bajo estudio han sido entrenadas y evaluadas usando datos reales proporcionados por el

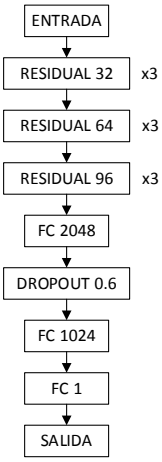


Fig. 3: Esquema de la red convolucional propuesta.

Departamento de Transporte de California a través de la plataforma PeMS. Para entrenar y validar la red, se han utilizado datos correspondientes a intensidad del tráfico (expresada en vehículos por unidad de tiempo), velocidad media temporal del tráfico (expresada en millas por hora), y densidad del tráfico (expresada en forma de porcentaje de ocupación de la vía).

El entrenamiento y evaluación de las redes ha sido realizado con datos correspondientes a la autopista I5. Esta autopista ha sido elegida por su longitud (cruza el estado de California de norte a sur, siendo la autopista más larga del estado), y la gran disponibilidad de sensores y datos. Para el presente estudio, se ha seleccionado un tramo de la I5 que va desde la frontera con México hasta pasada la población de Oceanside, y tiene una longitud de 89.67 millas (o 144.31 km). Se utilizan datos de 31 sensores desplegados en el tramo seleccionado; la Figura 4 muestra su localización en el tramo. Estos sensores se encuentran a una distancia media de 1.86 millas (o 2.99 km), y han sido elegidos por estar aproximadamente equidistantes.

La razón de buscar esta equidistancia es que las redes convolucionales comparten sus parámetros entrenables espacialmente, por tanto estos parámetros de la red serán iguales para los distintos sensores seleccionados, por lo que es preferible que estos sensores abarquen regiones de longitudes similares, de manera que la evolución del tráfico esté representada de la misma forma en cada sensor. Asegurando esta equidistancia se consigue que los píxeles de la imagen del tráfico resultante representen lo mismo.

Para cada sensor se han descargado los datos de tráfico (intensidad, velocidad media temporal, y densidad) con una granularidad temporal de 5 minutos para los años 2015 y 2016. En los datos originales del PeMS, la intensidad del tráfico es proporcionada en términos de vehículos cada cinco minutos y como suma de la intensidad de todos los carriles. Este valor de la intensidad se puede convertir a vehículos por hora y por carril siguiendo la ecuación (6):

$$Q' = \frac{Q}{n_{carriles}} \cdot 12 \qquad (6)$$

Donde Q es la intensidad en vehículos cada 5 minutos, Q' es la intensidad en vehículos por hora y por carril, y $n_{carriles}$ es el número de carriles del tramo en cuestión.

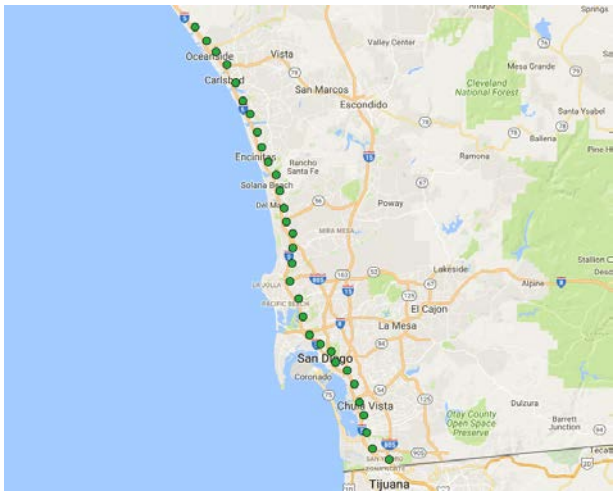


Fig. 4: Localización de los sensores de tráfico seleccionados (autopista I5, California, Estados Unidos).

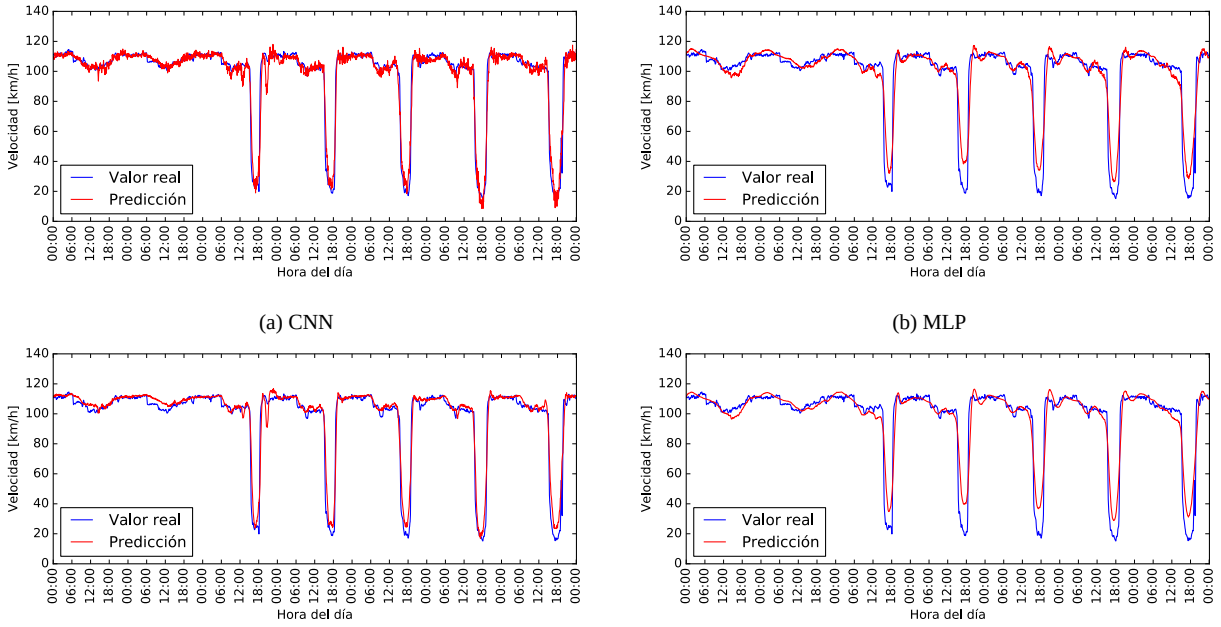
V. Resultados

El conjunto de datos descrito en la sección IV ha sido dividido en tres conjuntos, un conjunto de entrenamiento con el que entrenar los parámetros de la red, un conjunto de validación con el que monitorizar la precisión de la red durante el entrenamiento y decidir cuándo finalizar el proceso de entrenamiento, y un conjunto de test con el que evaluar la precisión final de la red. Todos los datos correspondientes al año 2015 han sido utilizados para confeccionar el conjunto de entrenamiento, mientras que la mitad de los datos de 2016 han sido utilizados para el conjunto de validación y la otra mitad para el de test.

Además de la red neuronal convolucional, también se ha entrenado un perceptrón multicapa sobre el mismo conjunto de datos para comparar los resultados de la red propuesta con los resultados de una red neuronal que sólo considera la evolución temporal del tráfico para predecir su estado futuro. La selección del perceptrón multicapa para la comparación se debe a dos razones: es la técnica para predicción de tráfico más utilizada en la literatura y su arquitectura es equivalente a la propuesta en este trabajo eliminando las capas convolucionales. El perceptrón multicapa entrenado tiene la misma arquitectura que las capas densas de la red convolucional, pues es el modelo que mejores resultados ha dado sobre los conjuntos de validación y test[2]. La red neuronal convolucional (CNN) y el perceptrón multicapa (MLP) han sido entrenados para predecir el valor de la velocidad media del tráfico en los siguientes 15 minutos utilizando como entrada los datos proporcionados por todos los sensores en el tramo bajo estudio durante las 6 horas previas al momento de predicción. De esta forma, la red CNN tiene como entrada una imagen formada por tres canales (uno por cada variable de tráfico: intensidad, velocidad media temporal y densidad) de 31 píxeles de alto y 72 píxeles de ancho (correspondientes a los 72 instantes anteriores con una granularidad de 5 minutos).

El rendimiento de las redes es evaluado mediante métricas de error que estiman la diferencia entre la velocidad media predicha y su valor real. En concreto,

[2]Se han probado arquitecturas de MLP con más capas y con diferente número de neuronas en cada capa.



(a) CNN

(b) MLP

(c) 10 CNN

(d) 10 MLP

Fig. 5: Predicción de la velocidad del tráfico de los distintos modelos y conjuntos de modelos para una semana.

se utilizan como métricas el error absoluto medio (MAE, *Mean Absolute Error*) y el error relativo o error porcentual absoluto medio (MAPE, *Mean Absolute Percentage Error*):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \tag{7}$$

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\frac{|\hat{y}_i - y_i|}{y_i} \tag{8}$$

Donde $\hat{y}_i$ es el valor predicho por la red, y_i es el valor real a predecir, y N es el número de instancias que forman la muestra sobre la que se calculan los errores.

La Tabla I compara el rendimiento de las dos redes objeto de estudio para los conjuntos de validación y test (resultados correspondientes a CNN y MLP). Los resultados muestran claramente que explotar la información espaciotemporal del tráfico a través de la red neuronal convolucional propuesta mejora de forma notable la precisión de la predicción del tráfico con respecto a considerar sólo su evolución temporal a través de un perceptrón multicapa. En concreto, la red convolucional propuesta mejora en un 22% el error absoluto medio, y en un 35% el error relativo. La tabla también muestra los resultados obtenidos por un conjunto de diez redes convolucionales (10 CNN) y un conjunto de 10 perceptrones multicapa (10 MLP). Cada conjunto está formado por diez redes neuronales entrenadas partiendo de parámetros entrenables aleatorios, de manera que al finalizar el entrenamiento las diez redes son diferentes. La salida de estos conjuntos de redes neuronales es el promedio de la salida de las diez redes que lo forman. Los conjuntos de modelos tienen como principal propiedad la reducción de la varianza de la salida, lo que suele dar lugar a mejores resultados. La Tabla I muestra que, para este estudio, utilizar conjuntos de redes neuronales mejora el error absoluto medio, pero empeora el error relativo.

TABLA I: RESULTADOS EN LOS CONJUNTOS DE VALIDACIÓN Y TEST

	Validación		Test	
	MAE	MAPE	MAE	MAPE
CNN	**4.08**	**5.81%**	4.08	**5.77%**
10 CNN	-	-	**3.73**	5.88%
MLP	5.26	8.88%	5.29	8.88%
10 MLP	-	-	5.27	8.98%

Esto es debido a que al promediar la salida de las diez redes neuronales y reducir la varianza, la predicción se ajusta mejor cuando la evolución de la velocidad del tráfico es más estable y toma valores más altos. Este efecto tiene más peso sobre el MAE. Por el contrario, si la velocidad toma valores más bajos, como en situaciones de congestión, la predicción del conjunto tiende a predecir valores de velocidad más altos que los que realmente se dan. Este efecto tiene más peso sobre el MAPE. Como el conjunto de validación se utiliza para monitorizar el error cometido por la red durante el entrenamiento y los conjuntos no se generan hasta que todas las redes que los forman están entrenadas, no se evalúa el error en validación de los conjuntos. La Figura 5 muestra la predicción de la velocidad obtenida con ambos modelos (CNN y MLP) y ambos conjuntos de modelos (10 CNN y 10 MLP) durante una semana. La figura muestra claramente que la salida de la red convolucional se ajusta mejor a la evolución real del tráfico que la del perceptrón multicapa. Es destacable el buen comportamiento de la red convolucional en las horas punta, cuando la velocidad del tráfico toma valores muy bajos debido a la congestión de la carretera. Estas horas punta son los momentos más críticos de la evolución del tráfico y suponen los instantes en los que resulta más útil una predicción precisa de la velocidad, pues permite a los gestores de tráfico anticiparse a los problemas que puedan surgir en la carretera.

VI. Conclusiones

Este trabajo propone una nueva técnica para predicción de tráfico basada en redes neuronales convolucionales que procesa la información del tráfico teniendo en cuenta su dimensión espacial y temporal. El rendimiento de la técnica propuesta ha sido comparado con el de una red de tipo perceptrón multicapa que centra el procesado de la información de tráfico en su evolución temporal. Los resultados obtenidos muestran que la red convolucional propuesta mejora sensiblemente la precisión en la predicción del estado futuro del tráfico, resaltando así la utilidad de procesar la información del tráfico teniendo en cuenta tanto su evolución temporal como espacial. Las mejoras son particularmente notables en posibles estados de congestión vial, lo cual es particularmente útil para poder realizar una gestión dinámica y eficaz del tráfico. A la vista de los resultados quedan abiertas algunas líneas de investigación como estudiar el efecto que la profundidad de la red neuronal convolucional u otros hiperparámetros tienen sobre la precisión de la predicción, analizar el rendimiento de la arquitectura de red neuronal propuesta para predicción de otras variables de tráfico distintas de la velocidad, o el efecto de la distancia entre sensores sobre la precisión de la predicción.

Agradecimientos

El presente trabajo ha sido financiado en parte por el Ministerio de Economía y Competitividad (TEC2014-5716-R), el Programa del Sistema Nacional de Garantía Juvenil (PEJ-2014-P-00524) y la Generalitat Valenciana (Proyectos de I+D+i desarrollados por grupos de investigación emergentes, GV-Convocatoria 2017).

Referencias

[1] Comisión Europea, "Política de transportes de la UE," [En línea]. Available: https://europa.eu/european-union/topics/transport_es.

[2] A. Valdés González-Roldán, S. de la Rica, M. Gullón y J. Azcoiti, Ingeniería de tráfico, Madrid: Bellisco, 2008.

[3] Y. LeCun, L. Bottou, Y. Bengio y P. Haffner, "Gradient-Based Learning Applied to Document Recognition," *Proceedings of the IEEE*, vol. 86, nº 11, pp. 2278-2324, 1998.

[4] California Department of Transportation, "Performance Measurement System," [En línea]. Available: http://pems.dot.ca.gov/.

[5] E. I. Vlahogianni, J. C. Golias y M. G. Karlaftis, "Short-term Traffic Forecasting: Overview of Objectives and Methods," *Transport Reviews*, vol. 24, nº 5, pp. 533-557, 2004.

[6] E. I. Vlahogianni, M. G. Karlaftis y J. C. Golias, "Short-term Traffic Forecasting: Where we are and where we're going," *Transportation Research Part C*, vol. 43, nº 1, pp. 3-19, 2014.

[7] B. L. Smith y M. J. Demetsky, "Short-term Traffic Flow Prediction: Neural Network Approach," *Transportation Research Record*, nº 1453, pp. 98-104, 1994.

[8] C. de Fabritiis, R. Ragona y G. Valenti, "Traffic Estimation and Prediction Based on Real Time Floating Car Data," *11th International IEEE Conference on Intelligent Transportation Systems*, pp. 197-203, Octubre 2008.

[9] J. Rzeszótko y S. H. Nguyen, "Machine Learning for Traffic Prediction," *Fundamenta Informaticae*, vol. 119, nº 3-4, pp. 407-420, 2012.

[10] J. W. C. van Lint, S. P. Hoogendoorn y H. J. van Zuylen, "Freeway Travel Time Prediction with State-Space Neural Networks: Modeling State-Space Dynamics with Recurrent Neural Networks," *Transportation Research Record*, nº 1811, pp. 30-39, 2002.

[11] X. Ma, Z. Tao, Y. Wang, H. Yu y Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transportation Research Part C*, vol. 54, pp. 187-197, 2015.

[12] Y. Tian y L. Pan, "Predicting Short-Term Traffic Flow by Long Short-Term Memory Recurrent Neural Network," *2015 IEEE International Conference on Smart City/SocialCom/SustainCom*, pp. 153-158, Diciembre 2015.

[13] Y. Lv, Y. Duan, W. Kang, Z. Li y F.-Y. Wang, "Traffic Flow Prediction With Big Data: A Deep Learning Approach," *IEEE transactions on Intelligent Transportation Systems*, vol. 16, nº 2, pp. 865-873, 2015.

[14] S. Dunne y B. Ghosh, "Weather Adaptive Traffic Prediction Using Neurowavelet Models," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, nº 1, pp. 370-379, 2013.

[15] H. Yin, S. C. Wong, J. Xu y C. K. Wong, "Urban traffic flow prediction using a fuzzy-neural approach," *Transportation Research Part C*, vol. 10, nº 2, pp. 85-98, 2002.

[16] B. M. Williams y L. A. Hoel, "Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results," *Journal of Transportation Engineering*, vol. 129, nº 6, pp. 664-672, 2003.

[17] T. Otoshi, Y. Ohsita, M. Murata, Y. Takahashi, K. Ishibashi y K. Shiomoto, "Traffic Prediction for Dynamic Traffic Engineering Considering Traffic Variation," *Computer Networks*, vol. 85, pp. 36-50, 2015.

[18] L. Vanajakshi, S. C. Subramanian y R. Sivanandan, "Travel time prediction under heterogeneous traffic conditions using global positioning system data from buses," *IET Intelligent Transport Systems*, vol. 3, nº 1, pp. 1-9, Marzo 2009.

[19] E. Cipriani, S. Gori, L. Mannini y S. Brinchi, "A procedure for urban travel time forecast based on advanced traffic data: Case study of Rome," *2014 IEEE 17th International Conference on Intelligent Transportation Systems*, pp. 936-941, Octubre 2014.

[20] W. Qiao, A. Haghani y M. Hamedi, "Short-Term Travel Time Prediction Considering the Effects of Weather," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2308, pp. 61-72, 2012.

[21] P. Cai, Y. Wang, G. Lu, P. Chen, C. Ding y J. Sun, "A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting," *Transportation Research Part C*, vol. 62, pp. 21-34, 2016.

[22] A. Hofleitner, R. Herring, P. Abbeel y A. Bayen, "Learning the dynamics of arterial traffic from probe data using a dynamic bayesian network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, nº 4, pp. 1679-1693, 2012.

[23] X. Jin, Y. Zhang y D. Yao, "Simultaneously prediction of network traffic flow based on PCA-SVR," *International Symposium on Neural Networks*, pp. 1022-1031, 2007.

[24] Y. Zheng, Q. Liu, E. Chen, Y. Ge y J. L. Zhao, "Time Series Classification Using Multi-Channels Deep Convolutional Neural Networks," *Web-Age Information Management: 15th International Conference, WAIM 2014, Macau, China, June 16-18, 2014. Proceedings*, Springer International Publishing, 2014, pp. 298-310.

[25] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens y Z. Wojna, "Rethinking the inception architecture for computer vision," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2818-2826, 2016.

[26] V. Nair y G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807-814, 2010.

[27] K. Simonyan y A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[28] K. He, X. Zhang, S. Ren y J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016.

[29] S. Ioffe y C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *International Conference on Machine Learning*, pp. 448-456, 2015.

[30] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever y R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, nº 1, pp. 1929-1958, 2014.

[31] M. Abadi, et al., "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.