Deep Learning of Transferable MIMO Channel Modes for 6G V2X Communications

Lorenzo Cazzella, *Graduate Student Member, IEEE,* Dario Tagliaferri, *Member, IEEE,*
Marouan Mizmizi, *Member, IEEE,* Damiano Badini, Christian Mazzucco,
Matteo Matteucci, *Member, IEEE,* and Umberto Spagnolini, *Senior Member, IEEE*

Abstract—In the emerging high mobility Vehicle-to-Everything (V2X) communications using millimeter Wave (mmWave) and sub-THz, Multiple-Input Multiple-Output (MIMO) channel estimation is an extremely challenging task. At mmWaves/sub-THz frequencies, MIMO channels exhibit few leading paths in the space-time domain (i.e., directions or arrival/departure and delays). Algebraic Low-rank (LR) channel estimation exploits space-time channel sparsity through the computation of *position-dependent* MIMO channel eigenmodes leveraging recurrent training vehicle passages in the coverage cell. LR requires vehicles' geographical positions and tens to hundreds of training vehicles' passages for *each* position, leading to significant complexity and control signalling overhead. Here we design a DL-based LR channel estimation method to infer MIMO channel eigenmodes in V2X urban settings, starting from a single LS channel estimate and without needing vehicle's position information. Numerical results show that the proposed method attains comparable Mean Squared Error (MSE) performance as the position-based LR. Moreover, we show that the proposed model can be trained on a reference scenario and be effectively transferred to urban contexts with different space-time channel features, providing comparable MSE performance without an explicit transfer learning procedure. This result eases the deployment in arbitrary dense urban scenarios.

Index Terms—MIMO, Deep learning, Channel estimation, V2X, Millimeter-wave, sub-THz, 6G

I. INTRODUCTION

MILLIMETER Wave (mmWave) $(30 - 100 \text{ GHz})$ and sub-THz $(100 - 300 \text{ GHz})$ bands arose as the leading solution to overcome the bandwidth scarcity occurring in the sub-6 GHz EM spectrum, e.g., $0.41 - 7.125$ GHz in 5G New Radio (NR) Frequency Range 1 (FR1). In particular, mmWaves in the $24.25 - 52.6$ GHz range are designated for 5G NR FR2 [1], while sub-THz W- and D-bands will be the pillars of 6G paradigm by 2030, to accommodate the increasing capacity requirements such as for Vehicle-to-Everything (V2X)-enabled services [2]. By increasing the carrier frequency, the propagation is affected by an orders-of-magnitude increase in the path-loss, inducing coverage reduction in Non Line-Of-Sight (NLOS) scenarios and a *sparse* communication channel characterized by few significant paths in the Space-Time (ST) domain of Directions of Arrival/Departure (DoAs/DoDs) and delays [3]–[5]. In this regard, Multiple-Input Multiple-Output (MIMO) systems, enabled by reduced antenna footprints at mmWave and sub-THz, are used to counteract the path-loss by beamforming strategies at both Transmitter (Tx) and Receiver (Rx) [6].

In MIMO systems, channel knowledge is essential for designing the correct Tx and Rx precoding/combining. Channel estimation methods can be classified as non-parametric, such as Least Squares (LS) or Minimum Mean Square Error (MMSE) [7], or parametric, e.g., Compressive Sensing [8], MUltiple SIgnal Classification (MUSIC) [9], or Estimation of Signal Parameters via Rotational Invariance Technique (ESPRIT) [10], [11]. For non-parametric approaches, the goal is to estimate the complex coefficients of the channel. LS channel estimation is known to be inaccurate in low Signal-to-Noise Ratio (SNR) conditions and large MIMO settings when the number of unknowns scales with the number of antennas and the bandwidth. The MMSE method shows superior performance, although it requires the knowledge of the Channel Covariance Matrix (CCM) and the correlation matrix of the received signals. Considering the challenge in acquiring CCM and covariance matrix of the received signals, the LS method is preferred over MMSE in practical deployments. Indeed, legacy multi-carrier systems, such as Orthogonal Frequency Division Multiplexing/Multiple Access (OFDM/OFDMA) 5G NR FR2 radio interface systems, leverage LS MIMO channel estimation from known pilot sequences [12], [13]. Nevertheless, parametric methods explore the physical structure of the propagation channel and estimate these channel parameters, i.e., angles of arrival/departure, delays, Doppler, and gain of each path. The class of non-parametric methods, in general, requires more training overhead and shows robust performance to the antenna array structure and residual hardware impairments [7]. On the other hand, the class of parametric methods requires lower training overhead to perform channel estimation but guarantees acceptable performances only for a sparse channel model and perfect hardware calibration. Residual hardware impairments, which are typical of practical systems, can lead to detrimental performance as described in [14], [15].

An alternative to a parametric channel estimation is based on algebraic theory. Algebraic Low-Rank (LR) methods combine high accuracy with an inherent robustness to hardware impairments [16], [17]. LR operate on multiple pilot sequences transmitted from a single (or multiple) collaborative User Equipment (UE) and collected by a fixed Base Station (BS), where each pilot transmission shares the same DoAs, DoDs and delays, while single paths' fading amplitudes are assumed to vary according to the Doppler spectrum. Indeed, the ensemble of received pilot sequences are used to compute the spatial and temporal *modes* of the MIMO channel to filter new pilot signals to retrieve the LR channel estimate [17].

From an algebraic point of view, LR only requires the *stationarity of the ST channel eigenmodes*, and there is no need to explicitly estimate DoAs, DoDs and delays, as the channel modes are *unstructured*, resulting more robust to antenna calibration issues. The LR efficacy is proportional to the sparsity of the MIMO channel matrix. Early works on LR were targeted to sub-6 GHz systems [16], while more recent ones were tailored for mmWave and compared with CS [17]. LR channel estimation leverages hundreds or thousands consecutive transmissions from the same moving collaborative UE towards the BS, limiting the application to static or quasi-static communication scenarios.

In our previous work [18], we overcome this limitation by collecting the set of received pilot sequences on *recurrent vehicle passages* over the same geographical area, to ensure the same ST channel structure for each received sequence. The key idea is that roads constrain vehicles to have recurrent passages and thus the associated MIMO channels share similar ST channel structures over different vehicles, as depicted in Fig. 1. The LR channel modes are thus related to physical UEs' positions in the cell, and this is suitable for V2X systems. Still, the method presented in [18] requires the availability of a suitable number (tens to hundreds) of collaborative UEs, i.e., vehicles transmitting their position, for *each* position within a given coverage cell. When the number of cells grows, the complexity of a position-based LR method rapidly becomes overwhelming. Furthermore, LR requires the continuous exchange of UEs position information, imposing a non-negligible BS-UE signalling.

Deep-Learning (DL) is foreseen to play a pivotal role in 6G, complementing or even substituting standard tasks introduced by novel communication frameworks, as MIMO systems at mmWave/sub-THz frequencies or reconfigurable intelligent surfaces [19], increasing the adaptability of the communication system to the local conditions of the environment. DL learns complex tasks from data where model-based techniques fail or turn out to be sub-optimal, exploiting Deep Neural Networks (DNNs) [20]–[22]. Recently, many works have addressed the problem of MIMO channel estimation by means of DL models. A seminal work in this direction is [20], which introduces the application of DL techniques at the physical layer. A DNN with a convolutional denoiser was derived from the *learned denoising-based approximate message passing* algorithm [23]. Channel estimation for fast time-varying MIMO OFDM systems in mobility is based on convolutional long short-term memory NN in [24]. Exploiting the deep image prior framework [25], the work [26] proposes instead a massive MIMO channel estimation method through an untrained deep neural network. A comparison of different DNN architectures (fully-connected DNNs, CNNs and bidirectional LSTMs) is presented in [27] to assist channel estimation in MIMO-OFDM systems. The work [28] proposes a deep learning-based channel estimation scheme based on LSTM and MLP to solve error propagation issues in data pilot-aided (DPA) channel estimation, while MIMO-OFDM pilot design and downlink channel estimation based on deep learning are achieved in [29], which provides also a pilot reduction technique based on neural network pruning. Transfer Learning

(a)

(b)
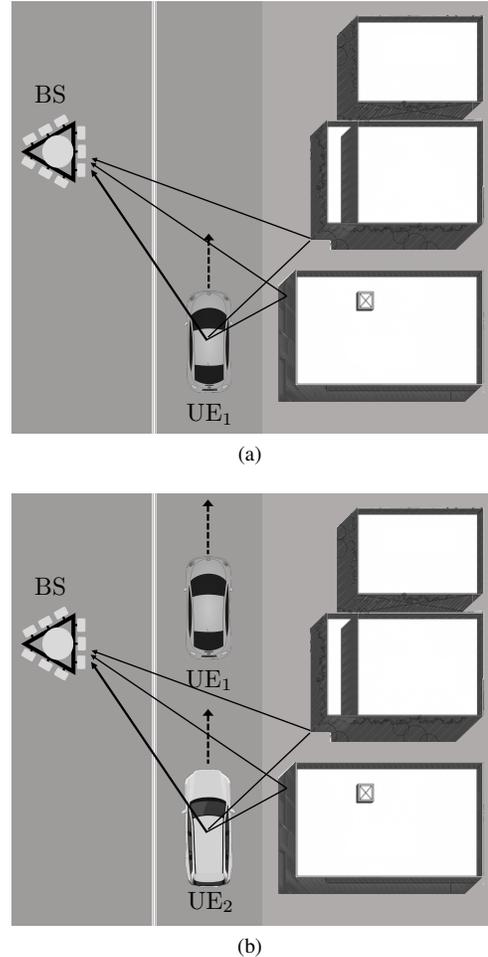
Fig. 1: Vehicular MIMO channel recurrences induced by road constraints: UE_1 and UE_2 experience the same DoDs, DoAs and delays in communicating with the BS when passing over the same location in the cell.

(TL) has been also recently considered as a powerful tool to extend and transfer the knowledge from one task to another that shares some inherent commonalities, by re-training only a subset of the DNN layers [30]. On channel estimation, a deep TL method exploits previously trained models to accelerate site adaptation [31]. The downlink channel prediction is addressed in [32] as a deep TL problem, proposing the use of fully-connected neural network architectures and fine-tuning trained models for new environments.

To advance with respect to current state of the art, in this paper we propose the following contributions:

- We propose a DL-based approach to infer the ST MIMO channel eigenmodes of LR channel estimation for 6G V2X with high-mobility. Compared to the reference position-based LR approach [18], advantages (after the initial training of the DNN) are the lack of UE position information at BS, and associated control signalling. Simulations by ray-tracing (to generate channel data) [33] over realistic vehicle trajectories [34] prove feasibility and benefits of the proposed DL-based LR approach, that outperforms the conventional LS estimation in terms of Mean Squared Error (MSE) by more than one order

of magnitude (≈ 15 dB on average). MSE performance of DL-based LR is comparable with position-based LR method for both frequency-flat and frequency-selective MIMO channels. Notice that MSE performance of the position-based LR method reaches the theoretical lower MSE bound in [16], thus our method is statistically efficient too.

- We show that the proposed DL-based LR channel estimation model generalizes over different urban scenarios, each characterized by different ST channel eigenmodes. Simulation results indicate that it is possible to perform the DNN training over a single scenario (exploiting pilot symbols from collaborative UEs in the reference position-based LR method) and transfer the learned algebraic MIMO channel structure to other scenarios, still outperforming LS. In particular, for frequency-flat MIMO channels, there is no practical advantage in employing additional TL procedures. For frequency-selective channels, the explicit re-training of the last 2 fully-connected layers of the DNN reduces the average MSE by ≈ 2 dB (consistently over 5 different scenarios).

We remark that the proposed approach is substantially different from the existing ones [23], [24], [26], which are targeted to learning either the physical (structured) channel features or directly the MIMO channel matrix entries. The advantage of the proposed approach is indeed the robustness against hardware impairments, inherited from the LR MIMO channel estimation [17]. Moreover, as shown in our previous work [35], MIMO channel eigenmodes can be effectively grouped in few (< 10) clusters in space, much less that the possible MIMO channel configurations. This characteristic eases the information transfer from one scenario to another, reducing the overall number of collaborative vehicular UEs used for DNN training to a single reference scenario. In this setting, the goal of the proposed DL-based LR estimation is not to outperform position-based LR, but rather to avoid the explicit signaling of the UE position for each pilot signal/LS channel estimate, exploiting the representational power of DL across different scenarios.

The paper is organized as follows: Section II outlines the analytical system and channel model used in this paper; Section III summarizes the reference position-based LR channel estimation method, functionally to the application of the DL technique in Section IV; Section V reports the simulation results while Section VI draws some final conclusions.

Notation

Bold upper- and lower-case letters describe matrices and column vectors. Matrix transposition, conjugation, conjugate transposition and Frobenius norm are indicated respectively as $\mathbf{A}^{\mathrm{T}}$, $\mathbf{A}^{*}$, $\mathbf{A}^{\mathrm{H}}$ and $\|\mathbf{A}\|$. $\mathrm{tr}(\mathbf{A})$, $\mathrm{rank}(\mathbf{A})$ extract, respectively, the trace and the rank of $\mathbf{A}$. Symbol $\otimes$ denotes the Kronecker product between two matrices. $\mathrm{vec}(\mathbf{A})$ denotes the vectorization by columns of $\mathbf{A}$. $\mathrm{diag}(\mathbf{A})$ denotes the extraction of the diagonal of $\mathbf{A}$, while $\mathrm{diag}(\mathbf{a})$ is the diagonal matrix given by vector $\mathbf{a}$. $\mathbf{I}_n$ is the identity matrix of size n. The Cholesky decomposition of a positive-definite matrix $\mathbf{A}$ is $\mathbf{A} = \mathbf{A}^{\frac{\mathrm{H}}{2}} \mathbf{A}^{\frac{1}{2}}$,

where $\mathbf{A}^{\frac{\mathrm{H}}{2}}$ is the lower-triangular unique square root of $\mathbf{A}$. The following property of the vectorization is used in the text: $\mathrm{vec}(\mathbf{AB}) = (\mathbf{B}^{\mathrm{T}} \otimes \mathbf{I}) \mathrm{vec}(\mathbf{A})$. With $\mathbf{a} \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{C})$ we denote a multi-variate circularly complex Gaussian random variable $\mathbf{a}$ with mean $\boldsymbol{\mu}$ and covariance $\mathbf{C}$. $\mathbb{E}[\cdot]$ is the expectation operator, while $\mathbb{R}$ and $\mathbb{C}$ stand for the set of real and complex numbers, respectively. δ_n is the Kronecker delta.

II. SYSTEM AND CHANNEL MODEL

We consider a single-user, multi-carrier uplink communication system over a bandwidth B, in which the Tx and the Rx are equipped with N_T and N_R antennas. At the receiving antennas, after the time and frequency synchronization and cyclic prefix removal, the Rx signal is:

$$\mathbf{y}(t) = \mathbf{H}(t) * \mathbf{x}(t) + \mathbf{n}(t), \tag{1}$$

where symbol $*$ denotes the matrix convolution between the transmitted signal $\mathbf{x}(t) = [x_1(t), \dots, x_{N_T}(t)]^{\mathrm{T}} \in \mathbb{C}^{N_T \times 1}$ at each Tx antenna and the $N_R N_T$ MIMO channel responses

$$\mathbf{H}(t) = \begin{bmatrix} h_{11}(t) & \cdots & h_{1N_T}(t) \\ h_{21}(t) & \cdots & h_{2N_T}(t) \\ \vdots & \vdots & \vdots \\ h_{N_R 1}(t) & \cdots & h_{N_R N_T}(t) \end{bmatrix} \in \mathbb{C}^{N_R \times N_T}, \tag{2}$$

where $h_{nm}(t)$ is the impulse response from the m-th Tx antenna to the n-th Rx antenna, whose maximum temporal support of the MIMO channel is limited to τ_{max}, $\forall n, m$. Vector $\mathbf{n}(t) \in \mathbb{C}^{N_R \times 1}$ denotes the additive Gaussian disturbance corrupting the received signal, comprising thermal noise and interference. By sampling (1) at time $t = wT$, where $T = 1/B$, we obtain the discrete-time signal

$$\mathbf{y}[w] = \mathbf{H}[w] * \mathbf{x}[w] + \mathbf{n}[w], \tag{3}$$

for $w = 0, \dots, W - 1$, where $W = \lceil \tau_{max}/T \rceil$ is the maximum number of channel taps and $\mathbf{H}[w] \equiv \mathbf{H}(wT)$ is the discrete-time MIMO channel matrix. For channel estimation purposes, the Tx signal $\mathbf{x}[w]$ is modelled as a random pilot sequence (known at the Rx), uncorrelated in time and space, i.e., $\mathbb{E}\left[\mathbf{x}[w]\mathbf{x}[\ell]^{\mathrm{H}}\right] = \sigma_x^2 \mathbf{I}_{N_T} \delta_{w-\ell}$ (σ_x^2 is the signal power). The w-th temporal sample of the noise at Rx array, $\mathbf{n}[w] = [n_1[w], ..., n_{N_R}[w]]^{\mathrm{T}} \in \mathbb{C}^{N_R \times 1}$, is modelled as a zero-mean circular Gaussian random vector with covariance matrix $\mathbf{Q}_n = \mathbb{E}\left[\mathbf{n}[w]\mathbf{n}[w]^{\mathrm{H}}\right]$, generally non diagonal (thus, correlated in space) due to the presence of a directional interference in the environment. We also assume that the noise is temporally uncorrelated, thus $\mathbb{E}\left[\mathbf{n}[w]\mathbf{n}[\ell]^{\mathrm{H}}\right] = \mathbf{Q}_n \delta_{w-\ell}$. Notice that the spatial covariance becomes diagonal, i.e., $\mathbf{Q}_n = \sigma_n^2 \mathbf{I}_{N_R}$, when either there is no interference in the environment (thermal noise only) or the interference can be assumed as isotropic [17]. In this latter setting, σ_n^2 denotes the noise power at each antenna. The SNR measured at each antenna is:

$$\mathrm{SNR} = \frac{\mathbb{E}\left[\left\| \sum_w \mathbf{H}[w] * \mathbf{x}[w] \right\|^2\right]}{\mathrm{tr}(\mathbf{Q}_n)}. \tag{4}$$

In the following, we detail the analytical model for the MIMO channel discrete impulse response $\mathbf{H}[w]$, to better

clarify the application of the DL-based LR channel estimation proposed in Section IV.

A. MIMO Channel Model

The mmWave/sub-THz MIMO channel impulse response is modelled as the sum of P paths as [3]:

$$\mathbf{H}(t) = \sum_{p=1}^{P} \beta_p\, e^{j2\pi\nu_p t}\, \mathbf{a}_R(\boldsymbol{\theta}_p)\mathbf{a}_T^{\mathrm{T}}(\boldsymbol{\phi}_p)\, g(t-\tau_p), \qquad (5)$$

where the p-th path amplitude β_p depends on path-loss and propagation geometry; ν_p is the p-th path Doppler shift; $\mathbf{a}_T(\boldsymbol{\phi}_p) \in \mathbb{C}^{N_T \times 1}$ and $\mathbf{a}_R(\boldsymbol{\theta}_p) \in \mathbb{C}^{N_R \times 1}$ are the Tx and Rx array response vectors to the p-th path, respectively, function of the DoDs $\boldsymbol{\phi}_p = [\phi_p^{\mathrm{az}}, \phi_p^{\mathrm{el}}]^{\mathrm{T}}$ and the DoAs $\boldsymbol{\theta}_p = [\theta_p^{\mathrm{az}}, \theta_p^{\mathrm{el}}]^{\mathrm{T}}$ (for azimuth and elevation); $g(t-\tau_p)$ is the pulse shaping waveform (typically a raised cosine) delayed by τ_p (p-th path delay). Without loss of generality, we consider half-wavelength spaced uniform planar arrays with isotropic antennas for both Tx and Rx. The Tx array response is structured as:

$$\mathbf{a}_T(\boldsymbol{\phi}_p) = \mathbf{a}_T^{\mathrm{el}}(\phi_p^{\mathrm{el}}) \otimes \mathbf{a}_T^{\mathrm{az}}(\phi_p^{\mathrm{az}}), \qquad (6)$$

where $\mathbf{a}_T^{\mathrm{az}}(\phi_p^{\mathrm{az}}) = [1, \ldots, e^{j\pi(N_T-1)\sin(\phi_p^{\mathrm{az}})}]$ and $\mathbf{a}_T^{\mathrm{el}}(\phi_p^{\mathrm{el}}) = [1, \ldots, e^{j\pi(N_T-1)\sin(\phi_p^{\mathrm{el}})}]$ are the steering vectors along azimuth and elevation DoDs. The Rx steering vector $\mathbf{a}_R(\boldsymbol{\theta}_p)$ is similarly structured. We also assume that the Doppler-related rotation is almost constant over $\tau_{\max}$ (normalized to the first echo), such that $\alpha_p = \beta_p\, e^{j2\pi\nu_p t} \sim \mathcal{CN}(0, \Omega_p)$, obeying the wide-sense stationary uncorrelated scattering model. The latter implies the uncorrelation between any pair of scattering amplitudes $\mathbb{E}\left[\alpha_{p,\ell}\alpha_{q,k}^*\right] = \Omega_p\delta_{p-q}\delta_{\ell-k}$, where $\alpha_{p,\ell}$ is the scattering amplitude of the p-th path of the ℓ-th channel.

By sampling (5) at $t = wT$ we obtain a compact matrix formulation of the MIMO channel

$$\mathbf{H}[w] = \sum_{p=1}^{P} \alpha_p\, \mathbf{a}_R(\boldsymbol{\theta}_p)\mathbf{a}_T^{\mathrm{T}}(\boldsymbol{\phi}_p)\, g[wT-\tau_p] =$$
$$= \mathbf{A}_R(\boldsymbol{\theta})\,\boldsymbol{\Lambda}[w]\mathbf{A}_T^{\mathrm{T}}(\boldsymbol{\phi}), \quad w = 0, \ldots, W-1 \qquad (7)$$

where $\mathbf{A}_T(\boldsymbol{\phi}) = [\mathbf{a}_T(\boldsymbol{\phi}_1), \ldots, \mathbf{a}_T(\boldsymbol{\phi}_P)] \in \mathbb{C}^{N_T \times P}$ and $\mathbf{A}_R(\boldsymbol{\theta}) = [\mathbf{a}_R(\boldsymbol{\theta}_1), \ldots, \mathbf{a}_R(\boldsymbol{\theta}_P)] \in \mathbb{C}^{N_R \times P}$ are two frequency-independent matrices embedding the spatial channel features, and $\boldsymbol{\Lambda}[w] = \mathrm{diag}(\alpha_1 g[wT-\tau_1], \ldots, \alpha_P g[wT-\tau_P]) \in \mathbb{C}^{P \times P}$ is a diagonal matrix collecting all the channel amplitudes scaled by the w-th tap of the pulse shaping waveform.

Algebraic analysis of the matrixes $\mathbf{A}_T(\boldsymbol{\phi})$ and $\mathbf{A}_R(\boldsymbol{\theta})$ defines the spatial diversity orders of the MIMO channel in terms of the number of distinguishable rays at Tx and Rx, given the number of antennas N_T and N_R. The diversity orders are expressed as

$$r_{\mathrm{S}}^{\mathrm{Tx}} = \mathrm{rank}(\mathbf{A}_T(\boldsymbol{\phi})) \le \min(N_T, P) \qquad (8)$$
$$r_{\mathrm{S}}^{\mathrm{Rx}} = \mathrm{rank}(\mathbf{A}_R(\boldsymbol{\theta})) \le \min(N_R, P) \qquad (9)$$

for Tx and Rx, respectively. From an algebraic point of view, $r_{\mathrm{S}}^{\mathrm{Tx}}$ and $r_{\mathrm{S}}^{\mathrm{Rx}}$ are the number of dimensions of the spatial subspaces spanned by the columns of $\mathbf{A}_T(\boldsymbol{\phi})$ and
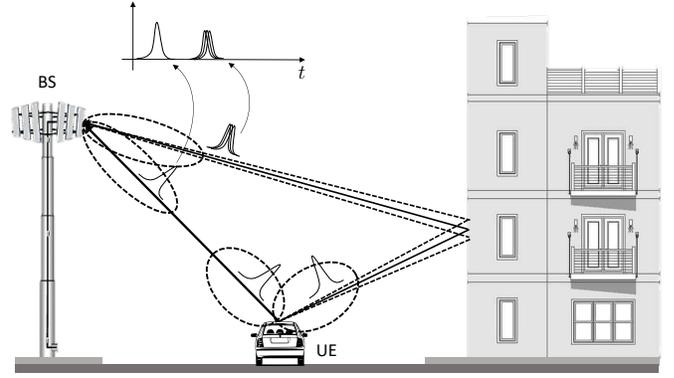
Fig. 2: Effect of Tx/Rx spatial and temporal selectivity on a multipath scenario ($P = 4$): reflections represented with dashed lines are spatially and temporally indistinguishable from the solid line one, due to the Tx and Rx beamwidths and bandwidth, therefore $r_{\mathrm{S}}^{\mathrm{Tx}} = r_{\mathrm{S}}^{\mathrm{Rx}} = r_{\mathrm{T}} = 2$.

$\mathbf{A}_R(\boldsymbol{\theta})$, respectively, i.e., the number of spatial eigenmodes representing the channel $\mathbf{H}[w]$. Orders are limited by either the number of channel paths or by the number of antennas. Usually, mmWave and sub-THz channels are characterized by $P < N_T, N_R$.

To ease the analytical derivations in Section III and the application of DL in Section IV, we can further manipulate (7) to extract the temporal (delays-related) diversity order of the MIMO channel as:

$$\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{A}}(\boldsymbol{\theta}, \boldsymbol{\phi})\, \mathbf{D}\, \mathbf{G}^{\mathrm{T}}(\boldsymbol{\tau}), \qquad (10)$$

where: $\boldsymbol{\mathcal{H}} = [\mathrm{vec}(\mathbf{H}[0]), \ldots, \mathrm{vec}(\mathbf{H}[W-1])] \in \mathbb{C}^{N_T N_R \times W}$ is the ST channel matrix, whose LS estimate is used as input to the DNN proposed in Section IV; $\boldsymbol{\mathcal{A}}(\boldsymbol{\theta}, \boldsymbol{\phi}) = [\mathbf{a}_T(\boldsymbol{\phi}_1) \otimes \mathbf{a}_R(\boldsymbol{\theta}_1), \ldots, \mathbf{a}_T(\boldsymbol{\phi}_P) \otimes \mathbf{a}_R(\boldsymbol{\theta}_P)]$ comprises both the DoDs and DoAs; $\mathbf{D} = \mathrm{diag}(\alpha_1, \ldots, \alpha_P)$, and matrix $\mathbf{G}(\boldsymbol{\tau}) = [\mathbf{g}(\tau_1), \ldots, \mathbf{g}(\tau_P)]$ embeds the temporal features $\boldsymbol{\tau} = [\tau_1, \ldots, \tau_P]$ through vectors $\mathbf{g}(\tau_p) \in \mathbb{R}^{W \times 1} = [g[-\tau_p], \ldots, g[(W-1)T-\tau_p]]^{\mathrm{T}}$. The temporal diversity order is therefore:

$$r_{\mathrm{T}} = \mathrm{rank}(\mathbf{G}(\boldsymbol{\tau})) \le \min(W, P). \qquad (11)$$

Similarly to (8)-(9), the temporal diversity order r_{T} has the meaning of number of temporally distinguishable paths of the MIMO channel, ruled by the temporal resolution of the system, i.e., the pulse width T (or equivalently by bandwidth B). For instance, for $B = 100$ MHz, a two-path channel with echoes spaced by 5 ns leads to $r_{\mathrm{T}} = 1$, as the temporal resolution of the system ($T = 1/B = 10$ ns) is not sufficient to distinguish each of the two paths. The algebraic interpretation of r_{T} is the number of dimensions of the temporal subspace of the channel, spanned by matrix $\mathbf{G}(\boldsymbol{\tau})$. The meaning of spatial and temporal channel orders is illustrated in Fig. 2, while the different channel manipulations used throughout the paper are reported in Table I. The knowledge of the channel diversity orders $r_{\mathrm{S}}^{\mathrm{Tx}}$, $r_{\mathrm{S}}^{\mathrm{Rx}}$ and r_{T} and of the spatial covariance of the noise allows to optimally estimate (in a statistical sense) the MIMO channel from multiple received pilot sequences sharing the same spatial

TABLE I: Channel manipulations

Symbol	Dimensions	Description
$\mathbf{h}$	$WN_T N_R \times 1$	time-space(Tx)-space(Rx) vector
$\mathcal{H}$	$N_T N_R \times W$	space(Tx+Rx)-time matrix
$\mathbf{H}[w]$	$N_R \times N_T$	space(Rx)-space(Tx) matrix (w-th sample)

and temporal channel subspaces (eigenmodes), as described in the next section.

III. POSITION-BASED LR CHANNEL ESTIMATION

This section reports the algebraic background for the LR channel estimation leveraging L different received pilot sequences $\{\mathbf{y}_\ell[w]\}_{\ell=1}^{\ell=L}$, assumed to be collected by the BS from *different* vehicular UEs passing in the same location $\bar{p}$ within the cell. Each UE is also requested to share with the BS its geographical position, obtained from on-board sensors or other techniques [36], [37]. Thus, sequences $\{\mathbf{y}_\ell[w]\}_{\ell=1}^{\ell=L}$ share the same ST propagation pattern, namely the same channel eigenmodes, while fading amplitudes can be arbitrarily varying for Doppler and mutually uncorrelated across the L pilot sequences. The complete analytical treatment, beyond the scope of the present work, can be found in [16]. In brief, the LR-estimated channel is retrieved through the application of a *pilot-specific* matrix $\mathbf{T}_\ell$, providing the conventional LS MIMO channel estimate, and a *position-specific* matrix $\mathbf{\Pi}_L(\bar{p})$ ($\bar{p}$ denotes a given position in the radio cell) on a single received pilot signal $\mathbf{y}_\ell = \left[\mathbf{y}_\ell^{\mathrm{T}}[0], \ldots, \mathbf{y}_\ell^{\mathrm{T}}[W-1]\right]^{\mathrm{T}} \in \mathbb{C}^{WN_R \times 1}$ collected from position $\bar{p}$:

$$\widehat{\mathbf{h}}_{LR,\ell} = \mathbf{\Pi}_L(\bar{p})\,\widehat{\mathbf{h}}_{LS,\ell}, \tag{12}$$

where $\widehat{\mathbf{h}}_{LR,\ell} \in \mathbb{C}^{WN_R N_T \times 1}$ is the LR-estimated channel vector and $\widehat{\mathbf{h}}_{LS,\ell} = \mathbf{T}_\ell\, \mathbf{y}_\ell \in \mathbb{C}^{WN_R N_T \times 1}$ is the conventional LS MIMO channel estimate, whose analytical expressions are detailed in [17]. Channel vector $\mathbf{h}$ (true or estimated) can be obtained from channel matrix $\mathcal{H}$ (true or estimated) by vectorization $\mathbf{h} = \mathrm{vec}(\mathcal{H})$.

The position-specific linear processing in (12) is designed in [16] as:

$$\mathbf{\Pi}_L(\bar{p}) = \mathbf{C}^{\frac{\mathrm{H}}{2}}\,\widehat{\mathbf{\Pi}}(\bar{p})\,\mathbf{C}^{-\frac{\mathrm{H}}{2}}, \tag{13}$$

where

- $\mathbf{C} \approx \frac{1}{\sigma_x^2}\left(\mathbf{I}_W \otimes \mathbf{I}_{N_T} \otimes \mathbf{Q}_n^{\mathrm{T}}\right)$ is the sample covariance matrix of the LS channel estimate $\widehat{\mathbf{h}}_{LS,\ell}$, needed to handle spatial/temporal noise correlations of interfering users in $\mathbf{Q}_n$;
- $\widehat{\mathbf{\Pi}}(\bar{p}) = \widehat{\mathbf{U}}(\bar{p})\widehat{\mathbf{U}}^{\mathrm{H}}(\bar{p})$ is the position-dependent projection matrix onto the ST propagation subspace associated to the ST basis (set of eigenmodes)

$$\widehat{\mathbf{U}}(\bar{p}) = \widehat{\mathbf{U}}_{\mathrm{T}}^* \otimes \widehat{\mathbf{U}}_{\mathrm{S}}^{\mathrm{Tx},*} \otimes \widehat{\mathbf{U}}_{\mathrm{S}}^{\mathrm{Rx}}. \tag{14}$$

Spatial ($\widehat{\mathbf{U}}_{\mathrm{S}}^{\mathrm{Tx}} \in \mathbb{C}^{N_T \times r_{\mathrm{S}}^{\mathrm{Tx}}}$, $\widehat{\mathbf{U}}_{\mathrm{S}}^{\mathrm{Rx}} \in \mathbb{C}^{N_R \times r_{\mathrm{S}}^{\mathrm{Rx}}}$) and temporal ($\widehat{\mathbf{U}}_{\mathrm{T}} \in \mathbb{C}^{W \times r_{\mathrm{T}}}$) MIMO channel eigenmodes are related to the set of DoDs, DoAs and delays, respectively. Eigenmodes form an orthonormal basis used to filter out from the LS estimate the noisy components that are not within the spanned algebraic subspace of the underlying channel. The eigenmodes $\widehat{\mathbf{U}}_{\mathrm{S}}^{\mathrm{Tx}}$, $\widehat{\mathbf{U}}_{\mathrm{S}}^{\mathrm{Rx}}$ and $\widehat{\mathbf{U}}_{\mathrm{T}}$ follow from the sample estimate of the spatial (Tx and Rx) and temporal correlation matrices, respectively

$$\widetilde{\mathbf{R}}_{\mathrm{S}}^{\mathrm{Tx}} = \frac{1}{L}\sum_{\ell=1}^{L}\sum_{w}\widetilde{\mathbf{H}}_{LS,\ell}[w]\,\widetilde{\mathbf{H}}_{LS,\ell}^{\mathrm{H}}[w], \tag{15}$$

$$\widetilde{\mathbf{R}}_{\mathrm{S}}^{\mathrm{Rx}} = \frac{1}{L}\sum_{\ell=1}^{L}\sum_{w}\widetilde{\mathbf{H}}_{LS,\ell}^{\mathrm{H}}[w]\,\widetilde{\mathbf{H}}_{LS,\ell}[w], \tag{16}$$

$$\widetilde{\mathbf{R}}_{\mathrm{T}} = \frac{1}{L}\sum_{\ell=1}^{L}\widetilde{\boldsymbol{\mathcal{H}}}_{LS,\ell}^{\mathrm{H}}\,\widetilde{\boldsymbol{\mathcal{H}}}_{LS,\ell}, \tag{17}$$

computed over L received pilot sequences from *different* vehicular UEs passing on *the same* position $\bar{p}$. In (15)-(17), $\widehat{\mathbf{H}}_{LS,\ell}^{\mathrm{H}}[w]$ and $\widehat{\boldsymbol{\mathcal{H}}}_{LS,\ell}$ are suitable arrangements (according to Table I) of the *whitened* channel $\widetilde{\mathbf{h}}_{LS,\ell} = \mathbf{C}^{-\frac{\mathrm{H}}{2}}\widehat{\mathbf{h}}_{LS,\ell}$. Spatial and temporal eigenmodes are then obtained as $\widehat{\mathbf{U}}_{\mathrm{S}}^{\mathrm{Tx}} = \mathrm{eig}_{r_{\mathrm{S}}^{\mathrm{Tx}}}(\widetilde{\mathbf{R}}_{\mathrm{S}}^{\mathrm{Tx}})$, $\widehat{\mathbf{U}}_{\mathrm{S}}^{\mathrm{Rx}} = \mathrm{eig}_{r_{\mathrm{S}}^{\mathrm{Rx}}}(\widetilde{\mathbf{R}}_{\mathrm{S}}^{\mathrm{Rx}})$ and $\widehat{\mathbf{U}}_{\mathrm{T}} = \mathrm{eig}_{r_{\mathrm{T}}}(\widetilde{\mathbf{R}}_{\mathrm{T}})$, i.e., from the $r_{\mathrm{S}}^{\mathrm{Tx}}$, $r_{\mathrm{S}}^{\mathrm{Rx}}$ and r_{T} leading eigenvectors of $\widetilde{\mathbf{R}}_{\mathrm{S}}^{\mathrm{Tx}}$, $\widetilde{\mathbf{R}}_{\mathrm{S}}^{\mathrm{Rx}}$ and $\widetilde{\mathbf{R}}_{\mathrm{T}}$. Notice that the directionality of the interference embedded in $\mathbf{C}$ (by means of matrix $\mathbf{Q}_n$) is typically estimated from the LS residual error [17]. Therefore, matrix $\mathbf{\Pi}_L(\bar{p})$ operates a position-based, noise-aware modal filtering on the standard LS MIMO channel estimate.

LR performance is proportional to the *sparsity degree* of the MIMO channel. It can be demonstrated that, if at least one of the following conditions holds [18]:

$$r_{\mathrm{S}}^{\mathrm{Tx}} < N_T, \quad r_{\mathrm{S}}^{\mathrm{Rx}} < N_R, \quad r_{\mathrm{T}} < W, \tag{18}$$

the LR method asymptotically ($L \to \infty$) outperforms LS. The value of L for practical convergence depends on N_T, N_R and W as well as on the SNR. For the MIMO settings and bandwidths considered in Section V, $L \approx 100$ guarantees the convergence, that is for each location of the coverage cell.

It is worth remarking that the application of LR does not require additional pilot signaling with respect to conventional communication systems. For instance, considering the 5G NR standard employing OFDM, LR can be enabled by the usage of the Demodulation Reference Signal (DM-RS). Further, no specific synchronization or cooperation among different UEs is needed. The key idea of LR is that it is sufficient to have enough channel samples for each position $\bar{p}$ (i.e., received pilot sequences/LS MIMO channel estimates) to guarantee the accurate estimation of the sample correlations (15), (16), (17). In this setting, the L pilot sequences from the same position $\bar{p}$ can be transmitted by different, uncoordinated UEs, provided that each sequence share the same spatial and temporal channel subspaces, i.e., the same channel eigenmodes, while the fading amplitudes can be mutually uncorrelated. Realistically, position $\bar{p}$ is the center of a geographical *region* with invariant ST channel features, whose spatial size is practically ruled by the accuracy of the positioning system. As shown in our previous work [18], it is enough to collect the channel samples with a position accuracy of $1 - 2$ m to reach the optimal (asymptotic) LR performance. For each region identified by $\bar{p}$, the BS can associate a list of eigenmodes $\{\widehat{\mathbf{U}}_{\mathrm{S}}^{\mathrm{Tx}}, \widehat{\mathbf{U}}_{\mathrm{S}}^{\mathrm{Rx}}, \widehat{\mathbf{U}}_{\mathrm{T}}\}$. The only additional signalling is the estimation of the UE's
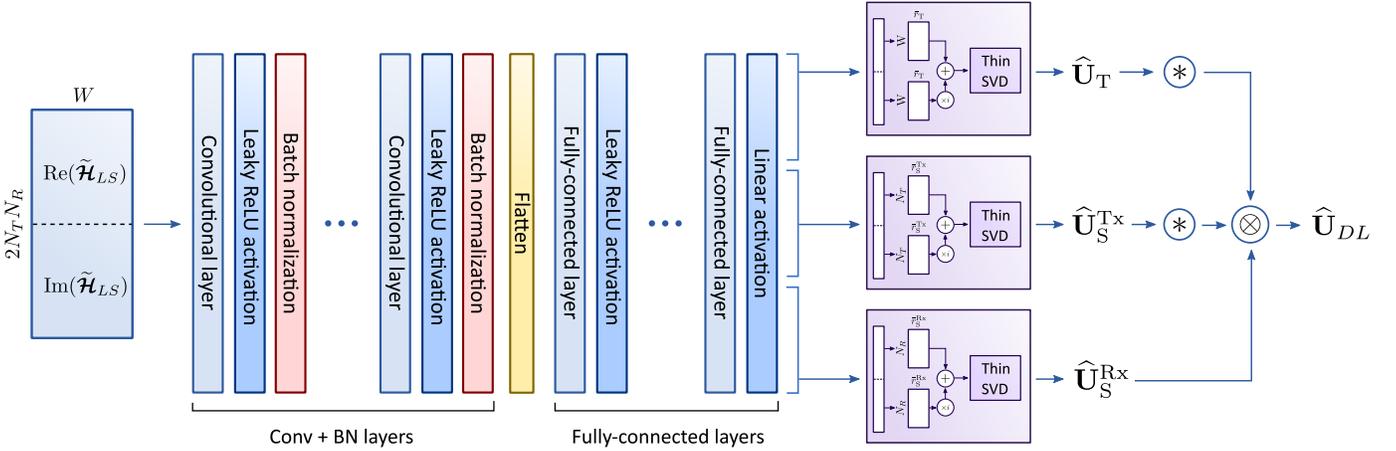
Fig. 3: Proposed DNN architecture. The network takes as input the real and imaginary parts of a LS channel estimate (after proper whitening by matrix $\mathbf{C}$), stacked on the spatial dimension, and outputs the corresponding Space-Time channel eigenmodes.

position $\widehat{p}$ for each transmitted pilot sequence. LR requires the knowledge of the UE position $\bar{p}$ during both the training phase (computation of $\mathbf{\Pi}_L(\bar{p})$) and the communication phase (run-time). The continuous exchange of position information in V2X systems is signalling intensive and increases the overhead on control channels. We explore in the following section a DL approach to retrieve the ST basis $\widehat{\mathbf{U}}(\bar{p})$ (and $\mathbf{\Pi}_L(\bar{p})$) directly from $\widehat{\mathbf{h}}_{LS,\ell}$, without the knowledge of the UE position.

IV. DL-BASED LR CHANNEL ESTIMATION

Leveraging the LR channel estimation algorithm described in Section III, we propose a DNN to infer the spatial and temporal LR MIMO channel eigenmodes $\mathbf{U}_T$, $\mathbf{U}_S^{\text{Tx}}$, and $\mathbf{U}_S^{\text{Rx}}$ from a single received pilot sequence, or, equivalently, a LS MIMO channel estimate. Exploiting the representational power of DL, we test its capability to capture recurring vehicular patterns in the neighborhood of the BS within an urban scenario, without requiring the explicit signalling of UE's position. In this regard, we use a large dataset of LS channel estimates $\{\widehat{\mathbf{h}}_{LS,m}\}_{m=1}^{m=M}$ gathered at the BS by multiple vehicular UEs along their path within the radio cell. The resulting labelled dataset $\{(\widehat{\mathbf{h}}_{LS,m}, \widehat{\mathbf{h}}_{LR,m})\}_{m=1}^{m=M}$ is composed by couples associating a given input LS channel estimate $\widehat{\mathbf{h}}_{LS,m}$ to the corresponding LR channel estimate $\widehat{\mathbf{h}}_{LR,m}$, provided by the position-based LR method in Section III. In this setting, M denotes the cardinality of the dataset, comprising a suitable number of different tracks covering the whole radio cell.

The proposed DNN is depicted in Fig. 3. An input LS-estimated ST channel matrix (after whitening) $\widetilde{\mathcal{H}}_{LS}$, is first normalized by the maximum absolute value of its elements and finally stacked by real and imaginary parts along the spatial dimension, leading to a $2N_T N_R \times W$ input matrix. We use a set of convolutional layers to extract effective features from the input channel matrix. Each convolutional layer employs the Leaky Rectified Linear Unit (Leaky ReLU) activation function [38]:

$$\Gamma(x) = \begin{cases} x & \text{for } x > 0 \\ 0.01x & \text{for } x \leq 0, \end{cases} \quad (19)$$

and is followed by a batch normalization layer [39], which speeds up network convergence and improves stability. After flattening the output of the last batch normalization layer to a single vector of convolutional features, the latter are mapped through a set of fully-connected layers to six matrices representing (grouped in pairs) the real and imaginary parts of three complex-valued matrices with the same dimensions of $\widehat{\mathbf{U}}_S^{\text{Tx}}$, $\widehat{\mathbf{U}}_S^{\text{Rx}}$ and $\widehat{\mathbf{U}}_T$ in (14), respectively. In order to output complex-valued unitary matrix representations, we project the three aforementioned complex-valued matrices on the corresponding Stiefel manifolds by applying thin Singular Value Decomposition (SVD$_{\text{th}}$), which is an efficient operator to diagonalize LR matrices [40]. SVD$_{\text{th}}$ decomposes a matrix $\mathbf{A} \in \mathbb{C}^{n \times r}$, with $r \leq n$, as:

$$[\mathbf{U}, \mathbf{s}, \mathbf{V}] = \text{SVD}_{\text{th}}(\mathbf{A}) \;\rightarrow\; \mathbf{A} = \mathbf{U} \, \text{diag}(\mathbf{s}) \, \mathbf{V}^H, \quad (20)$$

where $\mathbf{U} \in \mathbb{C}^{n \times r}$, $\mathbf{s} \in \mathbb{C}^{r \times 1}$, and $\mathbf{V} \in \mathbb{C}^{r \times r}$, with $\mathbf{U}$ and $\mathbf{V}$ unitary matrices. From (20), we consider only the $\mathbf{U}$ output, which is orthonormal and has the same dimensions as the target LR modes (14). Therefore, network training is carried only over $\mathbf{U}$, without updating the weights related to $\mathbf{s}$ and $\mathbf{V}$. Details on the automatic differentiation of complex-valued SVD can be found in [41]. The DNN input-output relation is therefore described by the nonlinear parametric mapping

$$\widehat{\mathbf{U}}_{DL} = f_{\mathbf{\Theta}}(\widetilde{\mathcal{H}}_{LS}), \quad (21)$$

where $\mathbf{\Theta}$ represents the network parameters to be optimized during training and $\widehat{\mathbf{U}}_{DL} = \widehat{\mathbf{U}}_T^* \otimes \widehat{\mathbf{U}}_S^{\text{Tx},*} \otimes \widehat{\mathbf{U}}_S^{\text{Rx}}$ is the DNN-inferred set of ST eigenmodes, aggregating the separate spatial and temporal eigenmodes as in (14). The LR-estimated MIMO channel is inferred as

$$\widehat{\mathbf{h}}_{LR}^{\text{pred}} = \mathbf{\Pi}_{DL} \, \widehat{\mathbf{h}}_{LS}. \quad (22)$$

where $\mathbf{\Pi}_{DL} = \mathbf{C}^{\frac{H}{2}} \widehat{\mathbf{U}}_{DL} \widehat{\mathbf{U}}_{DL}^H \mathbf{C}^{-\frac{H}{2}}$ is the DL-estimated counterpart of the position-specific matrix $\mathbf{\Pi}_L(\bar{p})$ in (12). Notice that $\mathbf{\Pi}_{DL}$ is not explicitly position-dependent. The selected training loss function, to be minimized over the DNN

parameters $\boldsymbol{\Theta}$, is the sum of the MSEs between the inferred LR channel estimates and the training ones:

$$\mathcal{L} = \sum_{m'=1}^{M'} \|\widehat{\mathbf{h}}_{LR,m'}^{\text{train}} - \widehat{\mathbf{h}}_{LR,m'}^{\text{pred}}\|^2, \qquad (23)$$

where $M' < M$ is the cardinality of the training dataset, a portion of the full one, $\widehat{\mathbf{h}}_{LR,m'}^{\text{train}}$ is the m'-th point LR MIMO channel estimate used for training and $\widehat{\mathbf{h}}_{LR,m'}^{\text{pred}}$ is from (22). In the simulations of Section V, the DNN parameters are optimized using the Adam [42] optimizer, updating the network weights at mini-batches of 32 data points.

An exhaustive analysis of computational complexity of the proposed method is beyond the scope of this work. However, for uplink the DNN training and prediction are performed at the BS, where computational constraints are more relaxed with respect to execution over specialized UE hardware. Moreover, we observe that, even if suboptimal, the decomposition of the channel eigenmodes into separate temporal and spatial components at Tx and Rx shown in Fig. 3 highly reduces the network training and inference computational complexity with respect to considering joint Space-Time channel eigenmodes. Indeed, the estimation of the latter would require to introduce, before the application of SVD_{th}, a network layer of dimension $2N_T N_R W \bar{r}$, as opposed to the proposed one with dimension $2(N_T \bar{r}_S^{\text{Tx}}) + 2(N_R \bar{r}_S^{\text{Rx}}) + 2(W \bar{r}_T)$. The computational cost for the automatic differentiation of complex-valued SVD performed during training can be easily derived from the procedure described in [41]. We refer the reader to [43] for a thorough analysis of CNNs' computational complexity under time constraint for real-time applications in industrial and commercial scenarios.

It is worth underlining that, differently from the MIMO channel eigenmodes obtained from the position-based LR method (Section III), which have variable diversity orders in space, i.e., $\{r_{S,m}^{\text{Tx}}, r_{S,m}^{\text{Rx}}, r_{T,m}\}_{m=1}^{m=M}$, all the unitary matrices inferred by the DNN have fixed orders $\bar{r}_S^{\text{Tx}}, \bar{r}_S^{\text{Rx}}, \bar{r}_T$. Fixed orders are needed as the output layer of the DNN has fixed dimension, and this implies that each set of channel eigenmodes (spatial and temporal) lie on the same Stiefel manifolds [44]. Notice that, considering a single cell scenario, the optimal value of $\bar{r}_S^{\text{Tx}}, \bar{r}_S^{\text{Rx}}, \bar{r}_T$ should guarantee the best possible modal filtering provided by $\boldsymbol{\Pi}_{DL}$ over the *whole* scenario. In principle, this shall imply to select the largest orders over the scenarios: $\bar{r}_S^{\text{Tx}} = \max\{r_{S,m}^{\text{Tx}}\}_{m=1}^{m=M}$, $\bar{r}_S^{\text{Rx}} = \max\{r_{S,m}^{\text{Rx}}\}_{m=1}^{m=M}$, $\bar{r}_T = \max\{r_{T,m}\}_{m=1}^{m=M}$. In practice, however, the true diversity orders of the channel are difficult to be estimated at each trajectory point within the cell. Moreover, the orders should be selected to enable a proper model transfer between different scenarios. Therefore, we consider the diversity orders as network hyperparameters to be optimized.

V. SIMULATION RESULTS

In this section, we present numerical results proving the effectiveness of the proposed DL-based LR channel estimation. Five scenarios are selected for numerical testing from portions of the city of Milan. They are depicted in Fig. 5, representing typical urban road crossings characterized by
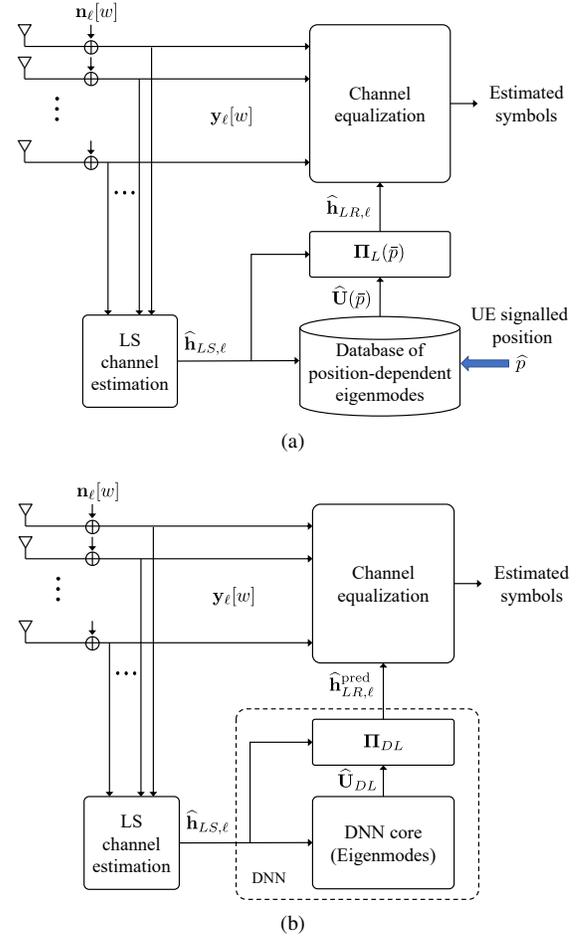
Fig. 4: Conceptual block scheme of the Rx implementing the position-based LR method (a) and the DL-based LR method (b). In (a), the estimated UE position $\widehat{p}$ is required to select the best eigenmodes from the database, while DL method (b) removes the position signalling constraint.

LOS propagation. Each scenario has a different geometry, road topology, and vehicular trajectory patterns, thus leading to diverse features in the ST domain. The simulation parameters are presented in Table II. We consider a OFDMA uplink communication at $f_0 = 28$ GHz carrier frequency between multiple vehicular UEs and a tri-sectoral BS, the former equipped with $N_T = 16$ (4×4) antennas and the latter with $N_R = 64$ (8×8) antennas (for each sector). The BS is located at 6 m from ground, in the position highlighted with a triangle in Fig. 5, while each UE moves at 1.5 m from ground. Two different communication bandwidths per UE are tested: $B = 1$ MHz, for which the channel is frequency-flat ($W = 1$), and 50 MHz, producing a frequency-selective channel ($W \gg 1$) in each of the five scenarios. We analyze the performance of the proposed channel estimation method in terms of Normalized Mean Squared Error (NMSE), defined as

$$\text{NMSE} = 10 \log_{10} \left(\frac{\mathbb{E}[\|\mathbf{h}_\ell - \widehat{\mathbf{h}}_{LR,\ell}\|^2]}{\mathbb{E}[\|\mathbf{h}_\ell - \widehat{\mathbf{h}}_{LS,\ell}\|^2]} \right), \qquad (24)$$

to highlight the MSE gain of LR compared to LS as reference method.
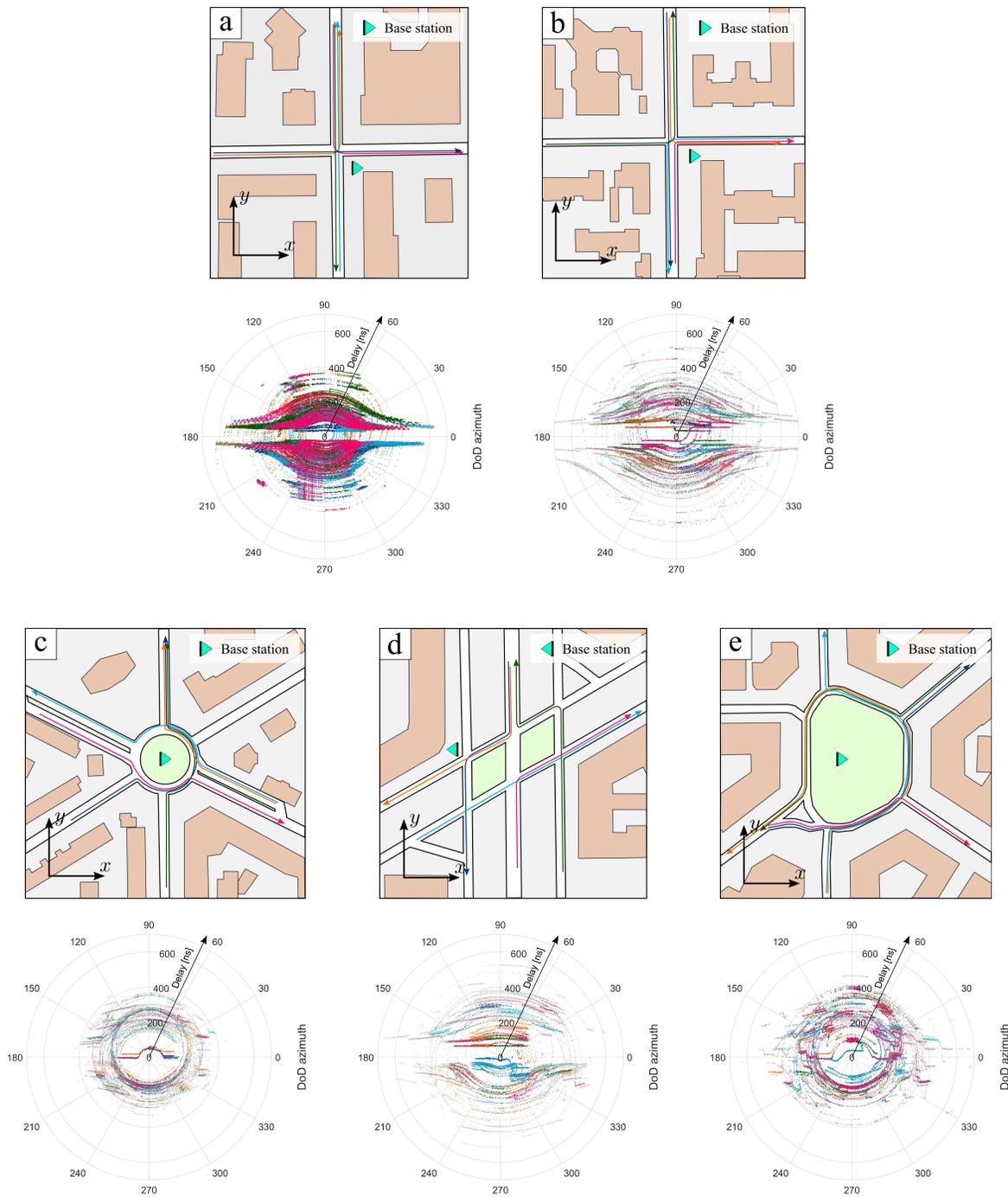
Fig. 5: Scenarios for the training and evaluation of the proposed DNN, with corresponding DoD/delay channel features (ray-tracing derived). Colored lines represent the reference vehicle trajectories, associated to colored paths in the DoD/delay plots.

A. Simulation setup

The datasets used for training the DNN over each scenario are produced by means of simulated channel data over realistic vehicle trajectories, obtained from SUMO (Simulation of Urban MObility) [34]. The mmWave channel parameters at 28 GHz are simulated by ray tracing using Altair WinProp [33] software, which provides for each considered geographical point the Direction of Departure (DoD) ψ, the Direction of Arrival (DoA) θ, the power Ω and the scattering amplitude α of each ray. The MIMO channel impulse response follows from

(7) by fixing the maximum number of taps over all the five scenarios ($W = 22$ for $B = 50$ MHz, determined by ray tracing). This enables the direct model transfer from one scenario to another. For each scenario, a dataset of $M = 2.5 \times 10^5$ channel samples has been produced considering multiple (different) realizations of 5 reference vehicular trajectories, where the Signal-to-Noise Ratio (SNR) has been fixed at 0 dB along all the trajectory.

We train the proposed model on scenario a), testing the learning capabilities of the DNN by comparing the NMSE of

TABLE II: Simulation parameters

Parameter	Symbol	Value
Carrier frequency	f_0	28 GHz
Bandwidth	B	1, 50 MHz
BS height from the ground	-	6 m
UEs height from the ground	-	1.5 m
Number of BS antennas	N_R	64 (8 × 8)
Number of UE antennas	N_T	16 (4 × 4)
Signal to Noise Ratio	SNR	0 dB

the DL-based LR method against the NMSE of the reference position-based LR introduced in Section III. Then, we analyse the generalization of the model to the remaining 4 scenarios b), c), d) and e) considering two distinct procedures: *(i)* testing the performance of the trained model by directly applying it to the new 4 scenarios, *without any retraining*; *(ii)* fine-tuning of the model trained on a) on the specific application scenario (b,c,d,e) by training only the last two fully-connected layers of the DNN (only for $B = 50$ MHz, since no improvement has been obtained by applying this procedure to the frequency-flat case). The DNN is trained using Adam optimizer [42] with a learning rate $\eta = 0.001$. Hyperparameters tuning is performed by means of grid search over the parameters presented in Table III (optimal selected hyperparameters are shown in bold).

B. Results for $B = 1$ MHz (frequency-flat)

We present the results obtained by applying the proposed DL-based channel estimation to frequency-flat MIMO channels with $W = 1$ temporal tap. After showing the NMSE performance of the model on the reference scenario a) (Fig. 5), we examine its generalization capabilities by directly applying it over scenarios b), c), d), e).

1) Performance of the DNN model on the reference urban scenario: The DNN model selected for frequency-flat channel estimation has 3 convolutional layers and 4 fully-connected layers. The first two convolutional layers use 64 filters while the last one uses a single filter. All the three convolutional layers use a 1×1 kernel, while the fully-connected layers are composed of 50 neurons each. This leads to $\approx 2.5 \times 10^5$ trainable model parameters $\boldsymbol{\Theta}$. We train our model with LS and LR channel estimates gathered on the reference scenario depicted in Fig. 5a). By hyperparameter search, we selected the ranks $\bar{r}_S^{\text{Tx}} = 4$, $\bar{r}_S^{\text{Rx}} = 8$, $\bar{r}_T = 1$ for the inferred unitary matrices corresponding to the spatial and temporal MIMO channel eigenmodes. Hence, with $\bar{r}_T = 1$, the MIMO channel is characterized by spatial modes only. The DL model converges within 10 training iterations to an average NMSE value of -14.9 dB (MSE gain with respect to LS estimation), to be compared with -15.7 dB NMSE provided by the position-based LR method in Section III. Figs. 6b and 6e show the NMSE performance of the proposed DL model when applied to 2 reference vehicular trajectories within the training scenario a). We consider multiple realizations of each trajectory to estimate pointwise the NMSE standard deviation for the inferred LR channel estimates (represented by the shaded gray area in Figs. 6b and 6e). The blue dashed line

is instead the mean NMSE provided by position-based LR channel estimation described in Section III (used for training), averaged over the whole length of the chosen trajectory. The results show that the DL-based NMSE closely matches the position-based NMSE except for some small performance penalty (< 2 dB). The same behavior has been also observed on the other 3 trajectory types over which the DNN model has been trained.

2) Generalization of the DNN model to different urban scenarios: To assess the effectiveness of the proposed DL method when challenged with new ST features of the environment, we test the model trained on the reference scenario a) against the b), c), d), and e) environments in Fig. 5. Notice that no transfer learning fine-tuning is used here. Our aim is to evaluate the capability of the model to map local convolutional features—learned from channel impulse responses sampled on the reference scenario—to the spatial and temporal MIMO channel eigenmodes on new data. Fig. 7a summarizes the NMSE of the channel estimates inferred over the tested scenarios by means of box plots, where the red line represents the median, the box encloses the interval between the first and the third quartiles, and the outer bars delimit the range of observed NMSE performances. We notice that the DL model transfer between one scenario to the others provides comparable NMSE performance, with only a slight increase of the NMSE dispersion. We also observed that TL fine-tuning does not provide any benefit, as the DL model is able to represent the MIMO spatial eigenmodes with the same accuracy experienced on a reference scenario. This result is particularly relevant for the implementation of the proposed DL-based channel estimation in practical systems, as it allows a remarkable reduction of the number of collaborative vehicles (UEs) used for training the DNN, at least for the frequency-flat channel case. In the considered settings, the DNN training dataset can be reduced by ≈ 80 %, as a full re-training of the DNN over the other 4 scenarios is not necessary.

C. Results for $B = 50$ MHz (frequency-selective)

We show the results obtained evaluating the proposed DL-based channel estimation method to frequency-selective MIMO channels, i.e., $B = 50$ MHz. After training the proposed model on the reference scenario a) in Fig. 5, we analyse—as for the previously discussed frequency-flat case—its generalization to scenarios b), c), d), e). In this case, we first test the direct application of the trained model to the new urban scenarios, without any further retraining, having fixed the temporal channel length $W = 22$ as the maximum over all the scenarios. We then examine whether any improvement can be obtained by retraining some network layers over LS and LR channel estimates proper of the specific application scenario.

1) Performance of the DNN model on the reference urban scenario: The considered DNN model has 3 convolutional layers and 4 fully-connected layers. The first two convolutional layers employ 64 filters, while the third one uses a single filter. Differently from the frequency-flat condition, the three convolutional layers use 1×3 convolutional kernels, in order to

TABLE III: Hyperparameter values considered in grid search for $B = 1$ MHz and $B = 50$ MHz.

Hyperparameters	$B = 1$ MHz	$B = 50$ MHz
Activation function for fully-connected layers	sigmoid, tanh, ReLU, **leaky-ReLU**	sigmoid, tanh, ReLU, **leaky-ReLU**
Activation function for convolutional layers	sigmoid, tanh, ReLU, **leaky-ReLU**	sigmoid, tanh, ReLU, **leaky-ReLU**
Number of units for fully-connected layers	20, **50**, 100, 200	20, 50, **100**, 200
Number of filters for convolutional layers	16, 32, **64**, 128	16, 32, **64**, 128
Kernel size for convolutional layers	**1x1**, 3x1, 5x1	1x1, **1x3**, 3x3, 1x5, 5x5
Rank for spatial channel eigenmodes at Tx ($\bar{r}_S^{Tx}$)	2, **4**, 6	2, **4**, 6
Rank for spatial channel eigenmodes at Rx ($\bar{r}_S^{Rx}$)	6, **8**, 10, 12	6, **8**, 10, 12
Rank for temporal channel eigenmodes ($\bar{r}_T$)	**1**, 2, 3	3, **5**, 7, 9
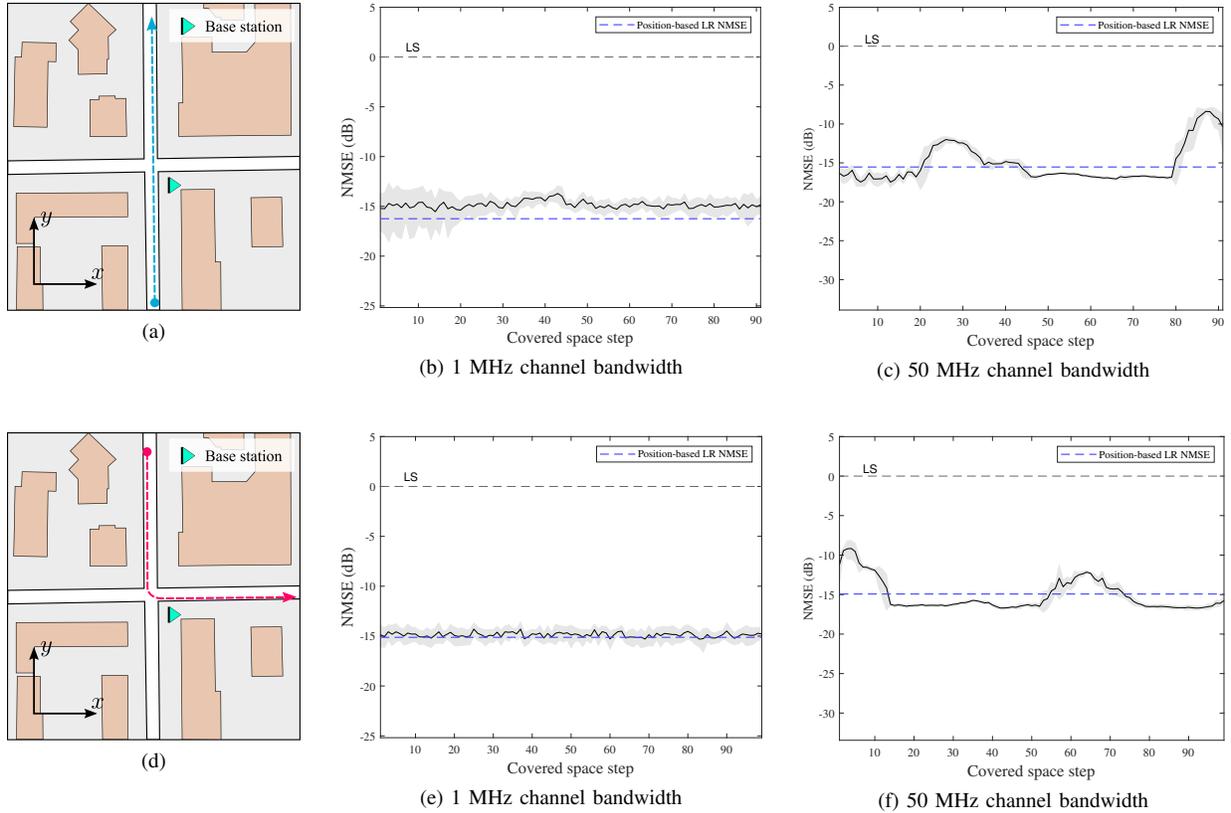
Fig. 6: Evaluation of the proposed method over sample trajectories in the reference scenario a) for $B = 1$ MHz and $B = 50$ MHz
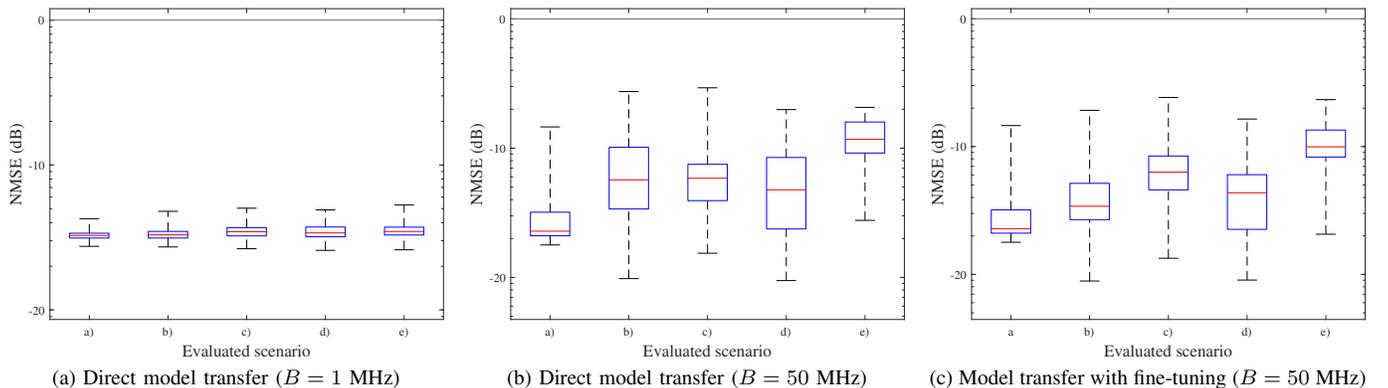
Fig. 7: Box plots showing median, interquartile range, and total range for the NMSE achieved using the DNN trained on scenario a) with direct model transfer over the remaining scenarios for $B = 1$ MHz (a), with direct model transfer over the remaining scenarios for $B = 50$ MHz (b), and with fine-tuning of the last two fully-connected layers on the specific scenario for $B = 50$ MHz (c).
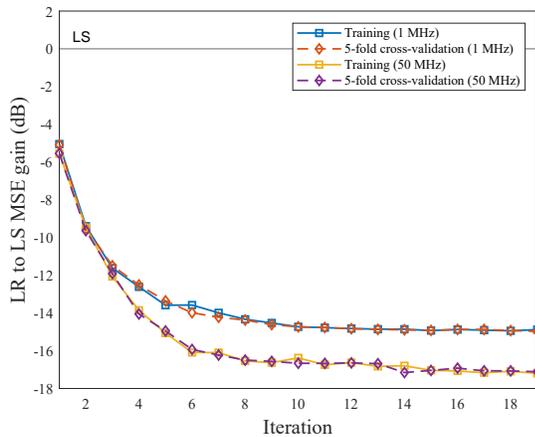
Fig. 8: NMSE gain achieved by the proposed DL method on the training set and on 5-fold cross-validation over training iterations.

jointly capture temporal features among consecutive temporal channel taps. The fully-connected layers are composed of 100 neurons each. This leads to $\approx 4.7 \times 10^6$ trainable model parameters Θ. We train the DNN model on the reference scenario a), selecting the diversity orders $\bar{r}_S^{Tx} = 4$, $\bar{r}_S^{Rx} = 8$, $\bar{r}_T = 5$. With the considered setting, the DNN converges within 10 training iterations to the average NMSE value of -16.7 dB, compared to the -17.1 dB obtained with the position-based LR method. Fig. 8 presents the NMSE gain attained by the proposed DL method on the training dataset and on 5-fold cross-validation over training iterations both at 1 MHz and 50 MHz bandwidths when scenario a) is considered. Training and cross-validation curves present the same NMSE gain behaviors for both the considered bandwidths, showing the generalization capabilities of the proposed architecture. We observed similar convergence behavior and generalization properties also for scenarios b), c), d), and e). Figs. 6c and 6f show the NMSE performance over the same 2 vehicular trajectories of, respectively, Figs. 6b and 6e (over scenario a)). Although with more variability, even in the $B = 50$ MHz case the DL model is able to provide comparable NMSE performance of the reference position-based LR method; as previously, a similar behavior is observed over the other three reference trajectories of a).

2) Generalization of the DNN model to different urban scenarios: To evaluate the generalization capabilities of the DL model in the frequency-selective channel case, we apply two different procedures: i) we directly test the model trained on reference scenario a) against scenarios b), c), d), and e), without any further re-training, and ii) starting from the model trained on scenario a), we fine-tune it by training only the last two fully-connected layers on the specific application scenario. We observed that the fine-tuning procedure converges after training the network with LS and LR channel estimates gathered in 10 vehicle passages for each trajectory type in the target urban scenario. Figures 7b and 7c show the NMSE performance achieved over the evaluated scenarios respectively for procedures i) and ii) by means of box plots, where the red line represents the median, the box encloses the interval between the first and the third quartiles, and the

outer bars delimit the range of observed NMSE performances. We notice that, compared to the frequency-flat case, the NMSE performance slightly deteriorate when transferring information to new scenarios, both in absence of retraining and with an explicit TL fine-tuning, still outperforming LS channel estimate by at least 10 dB. This effect is a consequence of a greater variability of the MIMO channel eigenmodes due to the non-negligible temporal component ($W \gg 1$). However, we did not observe any remarkable improvement applying a TL procedure, that only allows for a NMSE decrease of $1-2$ dB for scenarios b), d) and e).

VI. CONCLUSION

This paper addresses the problem of MIMO channel estimation in future 6G V2X systems proposing a novel DL-based LR channel estimation method. The proposed method leverages the received signal at the BS from road-induced recurrent vehicular UEs passages to design and train a DNN for the inference of MIMO channel eigenmodes. The goal is to improve conventional LS MIMO channel estimates without the need of any information on UEs' position. Exploiting the expressive power of DL and a training on LS and LR channel estimates collected over a whole radio cell, the proposed method requires only single input LS channel estimates to effectively infer the corresponding channel modes. Compared to a position-based LR channel estimation—which requires $L \approx 100$ pilot signals from as many vehicle passages for each location within a radio cell—this remarkably reduces any position-based training still achieving comparable NMSE performance.

Numerical results using realistic vehicular traffic and mmWave ray-tracing data show that the proposed DL-based LR method outperforms LS in terms of NMSE (≈ 15 dB) on channel estimation in both frequency-flat and frequency-selective channel cases, and attains the performance of the position-based LR, which in turn attains the theoretical MSE bound. Moreover, we show that the proposed DL model can be trained to infer the MIMO channel eigenmodes on a reference scenario, and then can be effectively transferred to urban scenarios (e.g., radio cells) characterized by substantially different space-time channel features, providing comparable NMSE performance without an explicit transfer learning fine-tuning procedure. This result allows to drastically reduce the number of training vehicles used to train the DNN, easing the practical implementation and motivating the application to future 6G V2X systems.

REFERENCES

[1] M. H. C. Garcia, A. Molina-Galan, M. Boban, J. Gozalvez, B. Coll-Perales, T. Şahin, and A. Kousaridas, "A tutorial on 5G NR V2X communications," *IEEE Communications Surveys Tutorials*, pp. 1–1, 2021.

[2] C. De Lima, D. Belot, R. Berkvens, A. Bourdoux, D. Dardari, M. Guillaud, M. Isomursu, E. S. Lohan, Y. Miao, A. N. Barreto, M. R. K. Aziz, J. Saloranta, T. Sanguanpuak, H. Sarieddeen, G. Seco-Granados, J. Suutala, T. Svensson, M. Valkama, B. Van Liempd, and H. Wymeersch, "Convergent communication, sensing and localization in 6G systems: An overview of technologies, opportunities and challenges," *IEEE Access*, vol. 9, pp. 26 902–26 925, 2021.

[3] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, 2014.

[4] C. Han, A. O. Bicen, and I. F. Akyildiz, "Multi-ray channel modeling and wideband characterization for wireless communications in the terahertz band," *IEEE Transactions on Wireless Communications*, vol. 14, no. 5, pp. 2402–2412, 2015.

[5] C.-X. Wang, J. Bian, J. Sun, W. Zhang, and M. Zhang, "A survey of 5G channel measurements and models," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3142–3168, 2018.

[6] S. Kutty and D. Sen, "Beamforming for millimeter wave communications: An inclusive survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 949–973, 2015.

[7] J. H. Emil Björnson and L. Sanguinetti, *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency*. Foundations and Trends in Signal Processing, 2017.

[8] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058–1076, 2010.

[9] Z. Guo, X. Wang, and W. Heng, "Millimeter-wave channel estimation based on 2-D beamspace music method," *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 5384–5394, 2017.

[10] A. Liao, Z. Gao, Y. Wu, H. Wang, and M. Alouini, "2D Unitary ESPRIT Based Super-Resolution Channel Estimation for Millimeter-Wave Massive MIMO With Hybrid Precoding," *IEEE Access*, vol. 5, pp. 24 747–24 757, 2017.

[11] M. Wang, F. Gao, S. Jin, and H. Lin, "An overview of enhanced massive mimo with array signal processing techniques," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 5, pp. 886–901, 2019.

[12] G. . V16.0.0, "3rd generation partnership project; technical specification nr; physical channels and modulation (release 16)," March 2020.

[13] 3GPP, "NR: Physical layer procedures for control," Third Generation Partnership Project (3GPP), Tech. Rep., 01 2020.

[14] R. Shafin, L. Liu, Y. Li, A. Wang, and J. Zhang, "Angle and delay estimation for 3-d massive mimo/fd-mimo systems based on parametric channel modeling," *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 5370–5383, 2017.

[15] D. Zhu, J. Choi, Q. Cheng, W. Xiao, and R. W. Heath, "High-resolution angle tracking for mobile wideband millimeter-wave systems with antenna array calibration," *IEEE Transactions on Wireless Communications*, vol. 17, no. 11, pp. 7173–7189, 2018.

[16] M. Nicoli, O. Simeone, and U. Spagnolini, "Multislot estimation of fast-varying space-time communication channels," *IEEE Transactions on Signal Processing*, vol. 51, no. 5, pp. 1184–1195, 2003.

[17] A. Brighente, M. Cerutti, M. Nicoli, S. Tomasin, and U. Spagnolini, "Estimation of Wideband Dynamic mmWave and THz Channels for 5G Systems and Beyond," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 9, pp. 2026–2040, 2020.

[18] M. Mizmizi, D. Tagliaferri, D. Badini, C. Mazzucco, and U. Spagnolini, "Channel Estimation for 6G V2X Hybrid Systems Using Multi-Vehicular Learning," *IEEE Access*, vol. 9, pp. 95 775–95 790, 2021.

[19] A. Salh, L. Audah, N. S. M. Shah, A. Alhammadi, Q. Abdullah, Y. H. Kim, S. A. Al-Gailani, S. A. Hamzah, B. A. F. Esmail, and A. A. Almohammedi, "A Survey on Deep Learning for Ultra-Reliable and Low-Latency Communications Challenges on 6G Wireless Systems," *IEEE Access*, vol. 9, pp. 55 098–55 131, 2021.

[20] T. O'Shea and J. Hoydis, "An Introduction to Deep Learning for the Physical Layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[21] C. Zhang, P. Patras, and H. Haddadi, "Deep Learning in Mobile and Wireless Networking: A Survey," *IEEE Communications Surveys Tutorials*, vol. 21, no. 3, pp. 2224–2287, thirdquarter 2019.

[22] H. Huang, S. Guo, G. Gui, Z. Yang, J. Zhang, H. Sari, and F. Adachi, "Deep Learning for Physical-Layer 5G Wireless Techniques: Opportunities, Challenges and Solutions," *IEEE Wireless Communications*, vol. 27, no. 1, pp. 214–222, Feb. 2020.

[23] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 748–751, 2018.

[24] Y. Liao, Y. Hua, and Y. Cai, "Deep learning based channel estimation algorithm for fast time-varying mimo-ofdm systems," *IEEE Communications Letters*, 2019.

[25] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9446–9454.

[26] E. Balevi, A. Doshi, and J. G. Andrews, "Massive mimo channel estimation with an untrained deep neural network," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 2079–2090, 2020.

[27] H. A. Le, T. Van Chien, T. H. Nguyen, H. Choo, V. D. Nguyen *et al.*, "Machine learning-based 5g-and-beyond channel estimation for mimo-ofdm communication systems," *Sensors*, vol. 21, no. 14, p. 4861, 2021.

[28] J. Pan, H. Shan, R. Li, Y. Wu, W. Wu, and T. Q. Quek, "Channel estimation based on deep learning in vehicle-to-everything environments," *IEEE Communications Letters*, vol. 25, no. 6, pp. 1891–1895, 2021.

[29] M. B. Mashhadi and D. Gündüz, "Pruning the pilots: Deep learning-based pilot design and channel estimation for mimo-ofdm systems," *IEEE Transactions on Wireless Communications*, vol. 20, no. 10, pp. 6315–6328, 2021.

[30] C. T. Nguyen, N. Van Huynh, N. H. Chu, Y. M. Saputra, D. T. Hoang, D. N. Nguyen, Q.-V. Pham, D. Niyato, E. Dutkiewicz, and W.-J. Hwang, "Transfer learning for future wireless networks: A comprehensive survey," *arXiv preprint arXiv:2102.07572*, 2021.

[31] W. Alves, I. Correa, N. González-Prelcic, and A. Klautau, "Deep transfer learning for site-specific channel estimation in low-resolution mmwave mimo," *IEEE Wireless Communications Letters*, 2021.

[32] Y. Yang, F. Gao, Z. Zhong, B. Ai, and A. Alkhateeb, "Deep transfer learning-based downlink channel prediction for FDD massive MIMO systems," *IEEE Transactions on Communications*, vol. 68, no. 12, pp. 7485–7497, 2020.

[33] "Altair WinProp," https://altairhyperworks.com/product/feko/winprop-propagation-modeling.

[34] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wießner, "Microscopic Traffic Simulation using SUMO," in *The 21st IEEE International Conference on Intelligent Transportation Systems*. IEEE, 2018. [Online]. Available: https://elib.dlr.de/124092/

[35] L. Cazzella, D. Tagliaferri, M. Mizmizi, M. Matteucci, D. Badini, C. Mazzuco, and U. Spagnolini, "Position-agnostic Algebraic Estimation of 6G V2X MIMO Channels via Unsupervised Learning," 2021.

[36] D. Tagliaferri, M. Brambilla, M. Nicoli, and U. Spagnolini, "Sensor-aided beamwidth and power control for next generation vehicular communications," *IEEE Access*, vol. 9, pp. 56 301–56 317, 2021.

[37] M. Mizmizi, S. Mandelli, S. Saur, and L. Reggiani, "Robust and flexible tracking of vehicles exploiting soft map-matching and data fusion," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, 2018, pp. 1–5.

[38] A. L. Maas, A. Y. Hannun, A. Y. Ng *et al.*, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. icml*, vol. 30, no. 1. Citeseer, 2013, p. 3.

[39] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. PMLR, 2015, pp. 448–456.

[40] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU press, 2012, vol. 3.

[41] Z.-Q. Wan and S.-X. Zhang, "Automatic differentiation for complex valued svd," *arXiv preprint arXiv:1909.02659*, 2019.

[42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[43] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 5353–5360.

[44] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

Lorenzo Cazzella is a PhD student in Information Technology at Politecnico di Milano. He received a B.Sc. degree (2017) in Information Engineering from Università del Salento and a M.Sc. degree (2020) in Computer Science and Engineering from Politecnico di Milano. During his master thesis, he has been involved in the Joint Lab between Huawei Technologies Italia and Politecnico di Milano, where his research work nowadays concerns the study and development of deep learning solutions for channel estimation and integrated communication and radar sensing at the infrastructure. His research interests are now focused on geometric deep learning and unsupervised learning techniques, with applications to wireless communication systems and radar systems.

Dario Tagliaferri (S'15-M'19) received the B.Sc. degree (2012), M.Sc. degree (2015) and the PhD (2019) in Telecommunication Engineering from Politecnico di Milano. He is currently assistant professor at Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Italy, in the framework of Huawei-Polimi Joint Research Lab. His research interests comprise signal processing techniques for navigation and wireless communication systems, in particular for sensor fusion of on-board sensors and MIMO channel estimation in 5G systems and beyond. He was the recipient of the Best Paper Award from the 1st IEEE International Online Symposium on Joint Communications & Sensing 2021.

Marouan Mizmizi (M'20) is a post-doctoral research fellow at Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Italy, in the framework of Huawei-Polimi Joint Research Lab. He received the B.Sc. degree (2012), M.Sc. degree (2015) in Telecommunication Engineering and the Ph.D. (2019) in Information Technology from Politecnico di Milano. His research interests comprise advanced signal processing techniques for V2X communication systems, in particular for MIMO channel estimation in mmWave and sub-THz frequencies and Relay selection and scheduling in high mobility scenarios. He was the recipient of the Best Paper Award from the 29th International Conference on Advanced Information Systems Engineering, 2017.

Damiano Badini joined the 5G Group of Huawei Technologies Italy Research Center in November 2016 as an Algorithm and System Engineer. His main areas of interest lie in mm-Wave technologies and communications. Before joining Huawei Technologies, he graduated cum laude from Politecnico di Milano in July 2013 with a Bachelor's degree in Electrical Engineering. From October 2013 to October 2014, he was Undergraduate Researcher at Nokia Bell Laboratories (Holmdel, NJ, USA) where he worked on the realization of the first real-time MIMO system in multimode optical fibers. In July 2016 he earned his master's degree with honors from Politecnico di Milano in Telecommunications Engineering.

Christian Mazzucco received the Laurea degree in telecommunications engineering from the University of Padova, Italy, in 2003, and the master's degree in information technology from the Politecnico di Milano, in 2004. In 2004, he joined Nokia Siemens Networks, Milan, where he was involved in research on UWB localization and tracking techniques. From 2005 to 2009, he was involved in several projects mainly researching and developing Wimax systems and high-speed LDPC decoders. In 2009, he joined Huawei Technologies, Italy, studying algorithms for high-power amplifiers digital predistortion, phase noise suppression, and MIMO for point-to-point microwave links. He is currently involved in researching phased array processing and the development of mmWave 5G BTS systems.

Matteo Matteucci (PhD) is Full Professor at Dipartimento di Elettronica Informazione e Bioingegneria of Politecnico di Milano, Italy. In 1999 he got a Laurea degree in Computer Engineering at Politecnico di Milano, in 2002 he got a Master of Science in Knowledge Discovery and Data Mining at Carnegie Mellon University (Pittsburgh, PA), and in 2003 he got a Ph.D. in Computer Engineering and Automation at Politecnico di Milano (Milan, Italy). His main research topics are pattern recognition, machine learning, machine perception, robotics, computer vision and signal processing. His main research interest is in developing, evaluating and applying, in a practical way, techniques for adaptation and learning to autonomous systems interacting with the physical world. He has co-authored more than 150 scientific international publications and he has been the principal investigator in national and international funded research projects on machine learning, autonomous robots, sensor fusion, and benchmarking of autonomous and intelligent systems.

Umberto Spagnolini is Professor of Statistical Signal Processing, Director of Joint Lab Huawei-Politecnico di Milano and Huawei Industry Chair. His research in statistical signal processing covers remote sensing and communication systems with more than 300 papers on peer-reviewed journals/conferences and patents. He is author of the book Statistical Signal Processing in Engineering (J. Wiley, 2017). The specific areas of interest include mmW channel estimation and space-time processing for single/multi-user wireless communication systems, cooperative and distributed inference methods including V2X systems, mmWave communication systems, parameter estimation/tracking, focusing and wavefield interpolation for remote sensing (UWB radar and oil exploration). He was recipient/co-recipient of Best Paper Awards on geophysical signal processing methods (from EAGE), array processing (ICASSP 2006) and distributed synchronization for wireless sensor networks (SPAWC 2007, WRECOM 2007). He is technical expert of standard-essential patents and IP. He served as part of IEEE Editorial boards as well as member in technical program committees of several conferences for all the areas of interests.