Corrections to “Computer Vision Aided mmWave Beam Alignment in V2X Communications”

Weihua Xu

***Abstract*—Due to some inappropriately adopted training steps for the deep neural networks (DNN) in the previous paper [1], the performance of both the proposed and compared beam alignment methods is underestimated in simulation. Thus, we slightly correct the design of the vehicle distribution feature (VDF) for the proposed vision based beam alignment when the MS location is available (VBALA). Specifically, we utilize the vehicle locations to expand the dimensions of the VDF. Then, we modify the training approach and correct the simulation results of Fig. 11, Fig. 12, Fig. 13, and Fig. 14 in [1]. All the conclusions derived from the simulation results remain unchanged.**

***Index Terms*—Computer vision, V2X communication, beam alignment, neural network, feature design**

I. Corrections to the VDF Design

Fig. 1 shows the corrected diagram of the proposed vision based beam alignment when the MS location is available (VBALA) [1]. We divide the $\mathrm{X_R}$-$\mathrm{Y_R}$ plane into G grids with length L_G and width W_G. Moreover, for the gth grid, $g = 1, 2, \cdots, G$, we set a local coordinate system (LCS) with $\mathrm{X_L}$-axis, $\mathrm{Y_L}$-axis, and $\mathrm{Z_L}$-axis, where the origin is the vertex $(i_g^\mathrm{X} W_\mathrm{G}, i_g^\mathrm{Y} L_\mathrm{G}, 0)$ and the $\mathrm{X_L}$-$\mathrm{Y_L}$-$\mathrm{Z_L}$ axis is parallel to the $\mathrm{X_R}$-$\mathrm{Y_R}$-$\mathrm{Z_R}$ axis. Then, for the vehicles whose plane locations are contained in the gth grid, we obtain their average length, width, height, and plane coordinates under the corresponding LCS as $l_{\mathrm{ave},g}$, $w_{\mathrm{ave},g}$, $h_{\mathrm{ave},g}$, and $(x_\mathrm{L}^g, y_\mathrm{L}^g)$, respectively.

The corrected vehicle distribution feature (VDF) is defined as a $G \times 6$ dimensional matrix $\boldsymbol{F}_\mathrm{cor}$, and the gth row of $\boldsymbol{F}_\mathrm{cor}$ is set as $[\frac{w_{\mathrm{ave},g}}{W_\mathrm{max}}, \frac{l_{\mathrm{ave},g}}{L_\mathrm{max}}, \frac{h_{\mathrm{ave},g}}{H_\mathrm{max}}, \frac{\theta_\mathrm{R}^g}{2\pi}, \frac{x_\mathrm{L}^g}{W_\mathrm{G}}, \frac{y_\mathrm{L}^g}{L_\mathrm{G}}]$. Compared with the original VDF $\boldsymbol{F}$ with $G \times 4$ dimensions in [1], $\boldsymbol{F}_\mathrm{cor}$ additionally utilizes the plane coordinates of the vehicles to expand each row. All other method steps remain unchanged.

II. Corrections to the Simulation Results

In the simulation of [1], the dataset for each deep neural networks (DNN) of beam alignment is only shuffled once at the initialization stage of the training phase, and there exists the statistical bias for the dataset. Thus, the performance of the proposed and compared methods is underestimated. We modify the training approach to the usual mode with dataset shuffle at every epoch. L_G and W_G are set as $6\mathrm{m}$ and $2\mathrm{m}$ respectively. The first 1D convolution layer of the original VDBAN is modified as 17 1D convolution layers. The filter numbers of the 17 1D convolution layers are 6, 6, 8, 8, 16, 16, 32, 32, 32, 64, 64, 128, 128, 256, 256, 512, and 512, respectively. The kernel sizes of the 17 1D convolution layers are 16, 16, 8, 8, 8, 8, 8, 8, 4, 4, 4, 4, 4, 4, 4, 2, and 1, respectively. The floating point operations (FLOPs) of the modified VDBAN is 2.26×10^8 and still significantly lower than the 1.22×10^{10} FLOPs of SIBAN. The grid size of the point cloud feature for LIDAR based beam alignment (LBA) is set as the same as VBALA. All other simulation settings are not changed.

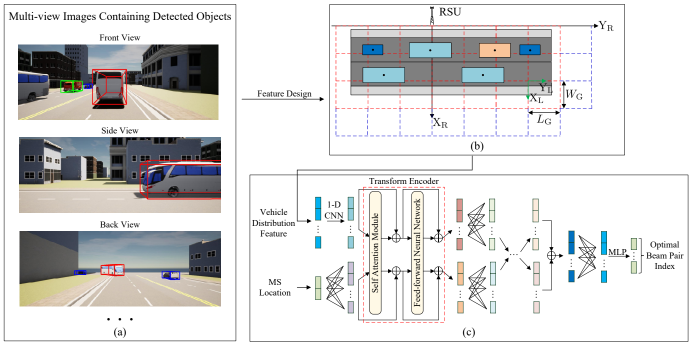

Fig. 1. The diagram of the proposed VBALA.

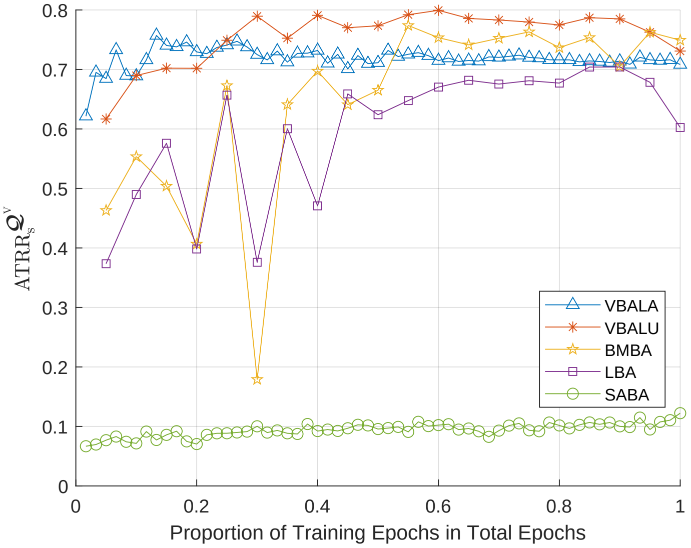

Fig. 2. $\mathrm{ATRR}_\mathrm{s}^{\mathfrak{Q}^\mathrm{V}}$ achieved by Top-1 beam pair selection with the increase of the number of training epochs.

We obtain the corrected simulation results in Fig. 2, Fig. 3, Fig. 4, and Fig. 5, which corresponds to the Fig. 11, Fig. 12, Fig. 13, and Fig. 14, respectively in [1]. As shown in Fig. 2, the DNNs of all the methods are trained to achieve the convergence. Due to the modified training approach, the $\mathrm{ATRR}_\mathrm{s}^{\mathfrak{Q}^\mathrm{V}}$ of all the methods are effectively improved compared with

W. Xu is with Institute for Artificial Intelligence, Tsinghua University (THUAI), Beijing National Research Center for Information Science and Technology (BNRist), Department of Automation, Tsinghua University, Beijing, P.R. China, 100084 (email: xwh19@mails.tsinghua.edu.cn).

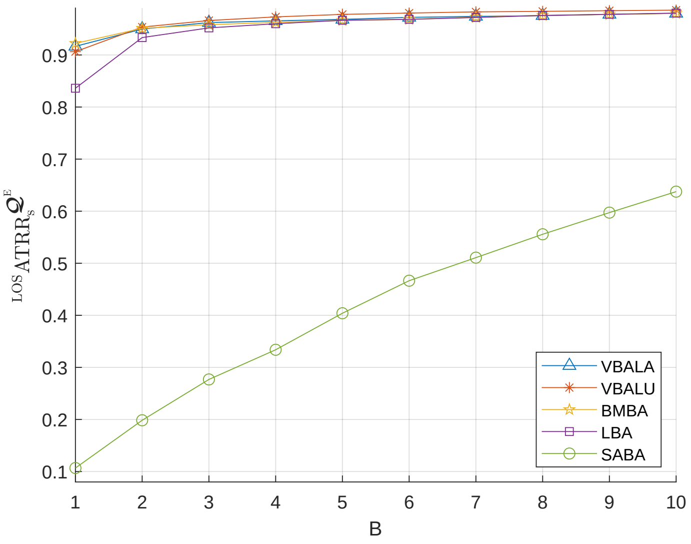

Fig. 3. ${}^{\mathrm{LOS}}\mathrm{ATRR}_{\mathrm{s}}^{\mathfrak{Q}^{\mathrm{E}}}$ for Top-B beam pair selection. The number of LOS test samples are 1930, which is 58% of total test samples.

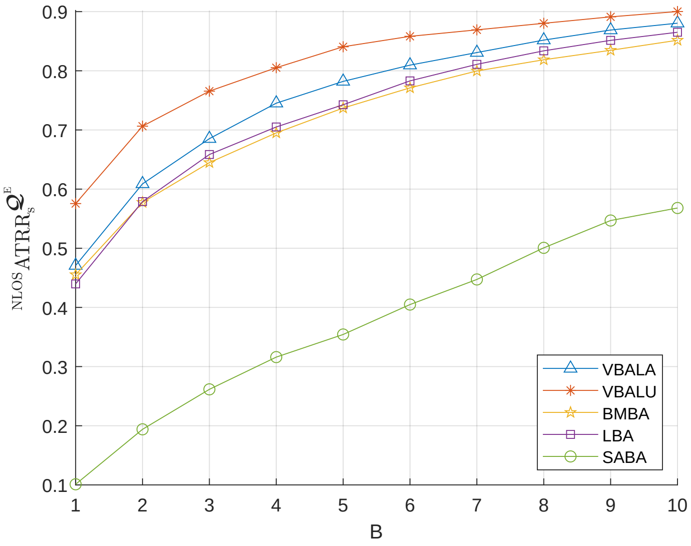

Fig. 4. ${}^{\mathrm{NLOS}}\mathrm{ATRR}_{\mathrm{s}}^{\mathfrak{Q}^{\mathrm{E}}}$ for Top-B beam pair selection. The number of NLOS test samples are 1410, which is 42% of total test samples.

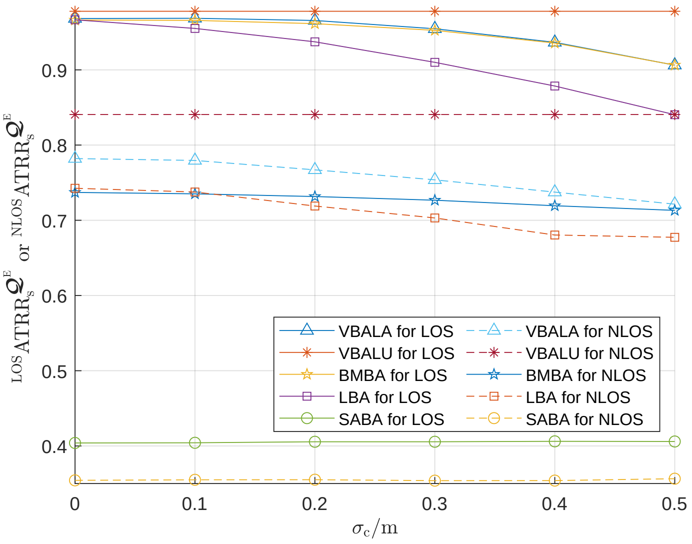

Fig. 5. ${}^{\mathrm{LOS}}\mathrm{ATRR}_{\mathrm{s}}^{\mathfrak{Q}^{\mathrm{E}}}$ and ${}^{\mathrm{NLOS}}\mathrm{ATRR}_{\mathrm{s}}^{\mathfrak{Q}^{\mathrm{E}}}$ for Top-5 beam pair selection with different location error $\mathcal{N}(0, \sigma_{\mathrm{c}}^2)$.

the Fig. 11 in [1]. All the conclusions from Fig. 3, the Fig. 4, and Fig. 5 are consistent with that from Fig. 12, Fig. 13, and Fig. 14 in [1]. The dataset and code for the proposed methods are publicly available [2].

References

[1] W. Xu, F. Gao, X. Tao, J. Zhang, and A. Alkhateeb, "Computer vision aided mmWave beam alignment in V2X communications," *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, pp. 2699–2714, Apr. 2023.

[2] https://github.com/whxuuuu/vision-communication-dataset.