Corrigendum and Addendum to "Computer Vision Aided mmWave Beam Alignment in V2X Communications"

Weihua Xu

***Abstract*—We are frightfully sorry for few defective details about the proposed vision based beam alignment method. Specifically, due to the limitations of the adopted training steps for the deep neural networks (DNN), the obtained simulation results of both the proposed and compared beam alignment methods do not fully demonstrate their achievable performance. Thus, we slightly correct the design of the vehicle distribution feature (VDF). The vehicle locations are used to expand the VDF. The utilized maximum vehicle size in each grid is replace by the average vehicle size. Then, we modify the training approach and add some simulation results to enhance the persuasiveness. All the conclusions remain unchanged.**

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

I. Corrigendum

Fig. 1 shows the diagram of the proposed vision based beam alignment when the MS location is available (VBALA) [1]. The maximum length $l_{\max,g}$, width $w_{\max,g}$ and height $h_{\max,g}$ of the vehicles in $\boldsymbol{\mathcal{V}}_g$ should be replaced with the average length $l_{\mathrm{ave},g}$, width $w_{\mathrm{ave},g}$ and height $h_{\mathrm{ave},g}$ of the vehicles in $\boldsymbol{\mathcal{V}}_g$. Moreover, for the gth grid, 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 $(i_g^{\mathrm{X}} L_{\mathrm{G}}, i_g^{\mathrm{Y}} W_{\mathrm{G}}, 0)$ 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. Thus, under the LCS of the gth grid, we obtain the average plane location coordinates of the vehicles in $\boldsymbol{\mathcal{V}}_g$ as $(x_{\mathrm{L}}^g, y_{\mathrm{L}}^g)$.

The VDF is defined as a $G \times 6$ dimensional matrix $\boldsymbol{F} \in \mathbb{R}^{G\times 6}$, and the gth row of $\boldsymbol{F}$ is set as $[\frac{l_{\mathrm{ave},g}}{L_{\max}}, \frac{w_{\mathrm{ave},g}}{W_{\max}}, \frac{h_{\mathrm{ave},g}}{H_{\max}}, \frac{\theta_{\mathrm{R}}^g}{2\pi}, \frac{x_{\mathrm{L}}^g}{L_{\mathrm{G}}}, \frac{y_{\mathrm{L}}^g}{W_{\mathrm{G}}}]$.

II. Addendum

In the simulation of [1], the dataset for each 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 $2\mathrm{m}$ and $6\mathrm{m}$ 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.

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 conver-

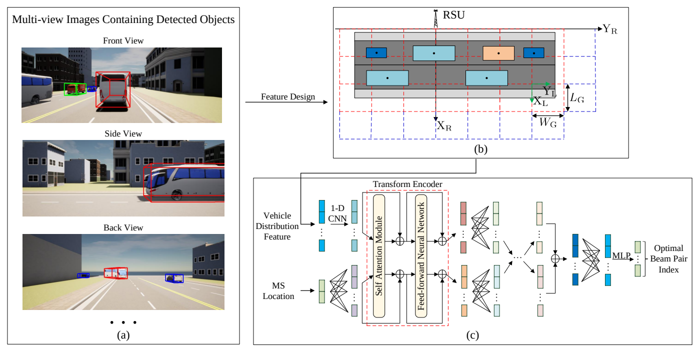

Fig. 1. The diagram of the proposed VBALA.

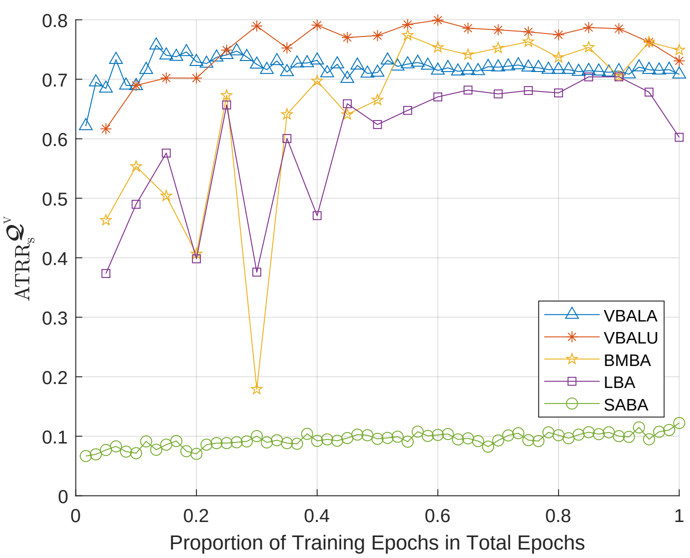

Fig. 2. $\mathrm{ATRR}_{\mathrm{s}}^{\mathfrak{Q}^{\mathrm{V}}}$ achieved by Top-1 beam pair selection with the increase of the number of training epochs.

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).

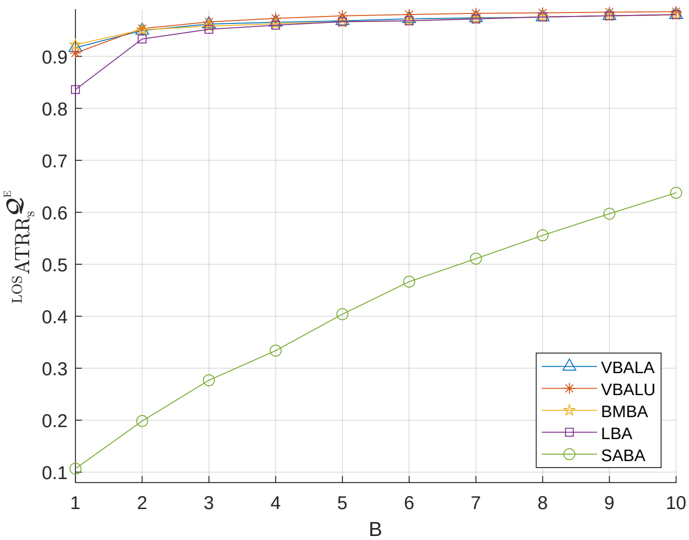

Fig. 3. $^{\mathrm{LOS}}\mathrm{ATRR}_{\mathrm{s}}^{\mathfrak{Q}^{\mathrm{E}}}$ for Top-B beam pair selection. The number of LOS test samples are 1930, which is 58% of total test samples.

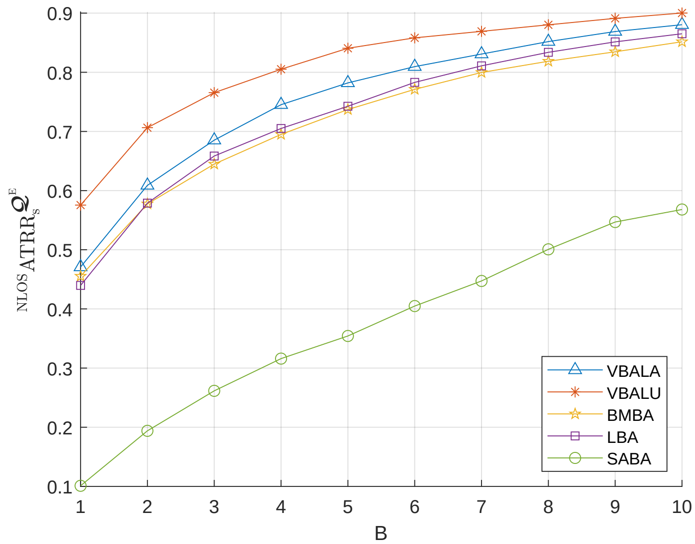

Fig. 4. $^{\mathrm{NLOS}}\mathrm{ATRR}_{\mathrm{s}}^{\mathfrak{Q}^{\mathrm{E}}}$ for Top-B beam pair selection. The number of NLOS test samples are 1410, which is 42% of total test samples.

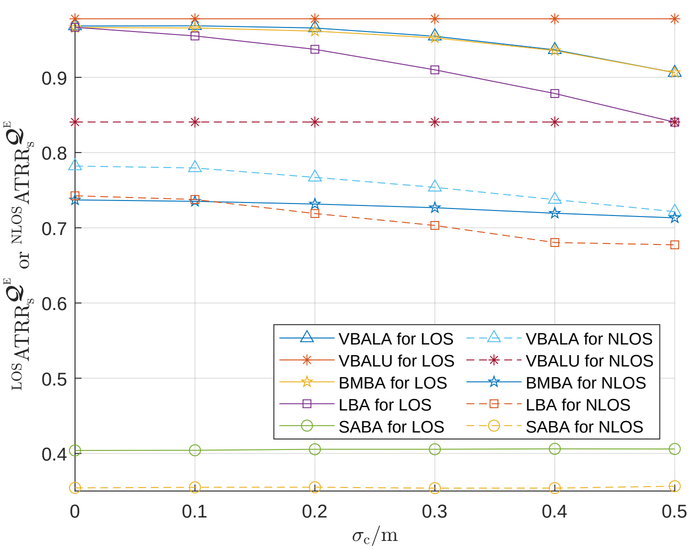

Fig. 5. $^{\mathrm{LOS}}\mathrm{ATRR}_{\mathrm{s}}^{\mathfrak{Q}^{\mathrm{E}}}$ and $^{\mathrm{NLOS}}\mathrm{ATRR}_{\mathrm{s}}^{\mathfrak{Q}^{\mathrm{E}}}$ for Top-5 beam pair selection with different location error $\mathcal{N}(0, \sigma_{\mathrm{c}}^2)$.

gence. Due to the modified training approach, the $\mathrm{ATRR}_{\mathrm{s}}^{\mathfrak{Q}^{\mathrm{V}}}$ of all the methods are effectively improved compared with 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.