Image Filtering Techniques for Beam Prediction in a Real-world 6G UAV Scenario

Vasileios P. Rekkas
ELEDIA@AUTH, School of Physics
Aristotle University of Thessaloniki
Thessaloniki, 541 24, Greece
Email: vrekkas@physics.auth.gr

Sotirios P. Sotiroudis
ELEDIA@AUTH, School of Physics
Aristotle University of Thessaloniki
Thessaloniki, 541 24, Greece
Email: ssoti@physics.auth.gr

Panagiotis Sarigiannids
Department of Informatics
and Telecommunications Engineering
University of Western Macedonia
Kozani, Greece
Email: psarigiannidis@uowm.gr

Konstantinos E. Psannis
Department of Applied Informatics
University of Macedonia
Thessaloniki, 541 24, Greece
kpsannis@uom.edu.gr

George K. Karagiannidis
School of Electrical and Computer Engineering
Aristotle University of Thessaloniki
Thessaloniki, 541 24, Greece
geokarag@auth.gr

Sotirios K. Goudos
ELEDIA@AUTH, School of Physics
Aristotle University of Thessaloniki
Thessaloniki, Greece
sgoudo@physics.auth.gr

2023 6th World Symposium on Communication Engineering (WSCE) | 979-8-3503-3950-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/WSCE59557.2023.10365979

***Abstract*—Millimeter-wave (mm-wave) and terahertz (THz) communication systems can satisfy the high data rate requirements in 5G, 6G, and beyond networks, but still rely on the use of extensive antenna arrays to guarantee sufficient received signal strength. Many antennas incur high beam training overhead; thus, the narrow beams require adjustment to support highly mobile applications. Deep learning (DL) vision-aided solutions can potentially forecast the optimal beams, leveraging raw RGB images captured at the base station. Image filtering techniques have been widely used in computer vision (CV) to modify and enhance the quality of an image, based on specific rules. This work applies different filters to RGB images for accurate mm-wave/THz beam prediction and feature extraction based on pre-trained convolutional neural networks (CNNs). The assessment of the developed framework is conducted on an actual dataset captured by an unmanned aerial vehicle (UAV) operating in the millimeter-wave (mm-wave) frequency range. The dataset comprises RGB images taken at the base station. Ensemble filtering techniques are also studied, enhancing the beam prediction accuracy of two state-of-the-art (SOTA) DL models.**

***Index Terms*—Millimeter wave, terahertz, deep learning, camera images, image filtering, beam prediction.**

I. Introduction

Vehicle-to-everything (V2X) communication has garnered significant research attention for both existing 5G and future 6G networks, catering to the needs of drones, unmanned aerial vehicles (UAVs), and autonomous vehicles [1]. To fulfill the high data rate demands of forthcoming aerial networks, unmanned aerial vehicles (UAVs) must be equipped with millimeter-wave (mm-Wave)/ terahertz (THz) transceivers and employ extensive antenna arrays. Also, by using narrow beams at both the transmitters and receivers, they ensure an adequate receive signal-to-noise ratio (SNR) [2].

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 957406 (TERMINET).

Beam-forming has been recognized as a crucial technique for mitigating signal attenuation in high-frequency bands, particularly in mm-Wave and THz communication systems [3]. The selection of optimal beams for large antenna arrays is often accompanied by a significant training overhead, motivating alternative research efforts to realize the rapid configuration of mm-Wave links.

Conventional methods for beam prediction involve the use of adaptive beam codebooks and the direct estimation of the beam-forming matrix from the channel matrix [4]. Although the beam training overhead is reduced, these approaches can only achieve a reduction in training overhead by approximately one order of magnitude.

Machine learning (ML) holds immense potential in driving significant advancements in 5G/B5G technology, offering automated methods for learning from spectrum data and executing intricate tasks such as spectrum sensing, classification, and waveform design. Deep learning (DL) rises as an emerging ML sub-field technology for 5G/ beyond-5G (B5G) technology, as it possesses an extraordinary capability to uncover intricate interconnections that are concealed within vast amounts of data. [5].

Authors in [6], [7] propose a novel ML approach, which utilizes supplementary sensory data, including visual and positional information, to achieve fast and precise beam prediction in mm-Wave/THz communication. For the vision-aided solution, various raw RGB images, from a real-world multi-modal UAV communication scenario, are utilized to address the beam prediction task. By leveraging the user's location within the visual scene, the problem can be addressed as a classification procedure. This approach involves assigning a beam index from a predefined codebook based on the user's position in the visual scene, with the operation frequency at 60 GHz.

In this study, various image filtering techniques are utilized

to improve the accuracy of beam prediction in two state-of-the-art (SOTA) deep learning (DL) models, namely Resnet-18 and VGG-16. As the given dataset is highly imbalanced, a resampling technique has been utilized to downsample the quantity of data from the majority classes to match the minority population, while classes with very few instances were excluded. Also, while authors in [6] downsample the beam powers of the 64 beams, to 32 elements to select as the optimal beam index, in this work all 64 beams are considered for the power vector that formulates the beam index. As a result, a corpus of 23 beam indexes with 3420 performances in total has been created. The main contributions of this work are summarized as follows:

- A database consisting of 23 beam indexes is generated using the power vector, which includes 64 beam powers.
- A comparative study of various image filters, is conducted to examine possible enhancement in SOTA DL models.
- An ensemble of image filters is applied to enhance accuracy results, with respect to the complexity.

To the best of the authors' knowledge, it is the first time that various image filters are applied and studied for a beam prediction classification problem in 6G networks, utilizing SOTA DL frameworks.

II. Data Processing

To enhance the performance of the SOTA DL architectures, various image pre-processing and filtering are employed to enhance the quality of the raw RGB images and map an image to a beam index, to perform this classification task.

A. Data Pre-processing

The initial dataset consists of 23 classes (i.e. beam indexes) and 3420 performances. To reduce the computational cost, the images are resized from a 960 × 540 format to a 64 × 64 format, while ensuring the preservation of essential information. Additionally, to enhance the efficiency of the computational process, the pixel values of the resized images are normalized.

B. Image Filtering

Image filtering is a technique employed to enhance the quality of an image by either removing specific features or accentuating other features present in the image [8]. In this work, a comparative study of image filtering techniques is conducted to evaluate the performance of Resnet-18 and VGG-16 for the classification and feature extraction task.

1) Gaussian Filtering: The Gaussian filtering method is based on peak detection and aims to treat peaks as impulses within the image. This filter effectively modifies both the spectral coefficient of interest and all the amplitude spectrum coefficients within its window. It acts as a linear low-pass filter, emphasizing pixels close to the edges, resulting in a reduction of edge blurring. The level of smoothing can be adjusted as desired and is computationally efficient [9].

2) Adaptive median filter: The median filter is a type of spatial filter that operates in a non-linear manner. By employing a designated window size, the median filter replaces the value of each pixel with the median value of the neighboring pixels in its vicinity. This filter effectively eliminates impulse noise from an image without causing any shifts in boundaries or reducing the overall contrast. However, the median filter impacts all pixels in the image, irrespective of their noise content. To address this limitation, adaptive median filters dynamically adjust the window size during the filtering process. The adaptive median filter operates on each pixel individually. It computes the median value using the standard median filtering process and then compares it with a predetermined threshold. Based on this comparison, the filter determines whether to replace the pixel, keep it unchanged, or increase the neighborhood size and recalculate. This method selectively applies the filter solely to image pixels that possess noise content, ensuring that the pixels unaffected by noise are preserved without any alterations [9].

3) Sharpening filter: Sharpening filters are employed to emphasize intricate details within an image, particularly the edges. Various spatial filters, such as the Laplacian, Laplacian of Gaussian, high-boost, and unsharp masking filters, along with their advanced variations, utilize linear convolution kernels to enhance image sharpness. Convolution is employed to alter the spatial properties of an image, manipulating the visual characteristics of the image through mathematical operations, leading to various outcomes such as smoothing, sharpening, enhancement, or emphasizing edges [10].

4) Box filter: The Box Filter is a type of low-pass filter used to smooth an image by uniformly averaging all samples within a square region of the image, removing details, noise, and edges from the image. The Box Filter operates by moving a window of a fixed (often square) size across the image. At each position, the filter determines the mean value of the pixel luminance or intensity within the window and replaces the central pixel with this computed average value. The dimensions of the window determine the level of smoothing or blurring that is applied to the image [11].

5) Non-local means (NLM) filter: The denoising of a pixel using the NLM filter involves calculating a weighted mean by considering all the pixels in the noisy image. These weights are determined based on the similarity of gray-level intensities between the local neighborhoods of the pixel being examined and the surrounding pixels that contribute to the mean. Essentially, the weights are directly proportional to the similarity between the two neighborhoods. When the neighborhoods are more alike, the corresponding weight is larger, resulting in a greater impact on the final denoised value [12].

6) Ensemble of filters: In this work, an ensemble filtering approach, combining box and sharpening filters, is proposed. Ensemble filtering is a strategy that involves merging various filters to amplify an image's quality by capitalizing on the distinct properties of each filter. Each filter contributes its unique features and capabilities to ultimately improve the output image. Ensemble models can train multiple models on

varied subsets of the data. This approach helps to enhance generalization and reduce the risk of overfitting. This diversity in training helps to improve the generalization capability of the ensemble, leading to more robust and reliable predictions.

By utilizing an ensemble of the box and sharpening filters, a more comprehensive image enhancement is achieved. By initially applying a box filter to effectively reduce noise and smooth the image, the subsequent application of a sharpening filter enhances edges and details, improving the prediction accuracy of the SOTA DL models.

III. Performance evaluation

This section presents the different experiments for the classification task of the beam indexes based on raw and filtered RGB images. First, the performance results of the SOTA pre-trained models are provided. Then, an analysis of the results for the proposed DL models, which are trained and validated on the dataset described in [6], is conducted. To study the generalization capability of the pre-trained CNNs, 5-fold cross-validation is utilized.

A. Performance metrics

ML and DL algorithms can be assessed based on their performance and ability to generalize using various performance and statistical metrics. When working within a supervised learning setup, various performance metrics can be utilized for multi-class classification tasks. The beam classification task is primarily evaluated using the accuracy metric. Accuracy can be defined as the portion or percentage of correct predictions out of the total predictions made, and can be mathematically expressed as:

$$\text{Accuracy} = \frac{1}{N}\sum_{n=N}^{N-1}\mathcal{F}(\hat{x}_n = x_n) \tag{1}$$

N is the number of test classes, $\hat{x}_n$ denotes the ML model class predicted label, x_n is the true class label, and $\mathcal{F}(x)$ represents the indicator function.

B. Experiments and Results

In this section, the performance results of the two SOTA pre-trained models are presented, along with a short analysis of the results.

1) DL Models: For the training of the models, the cross-entropy loss function is employed in conjunction with the Adam optimizer. The cross-entropy loss function is defined as:

$$\mathcal{L} = -\sum_{i=1}^{N} y_i \log(p_i) \tag{2}$$

$\mathcal{L}$ represents the cross-entropy loss function, N is the number of test classes, y_i denotes the true label for the ith sample, and p_i the predicted probability of the ith sample belonging to the corresponding class.

Each model is trained over 20 epochs, utilizing a mini-batch size of 4, whereas the learning rate is set to 3e-5. To avoid overfitting, the dropout technique is utilized, and the learning rate decay of the optimizer is set to 1e-4. Table I includes the hyperparameters that were utilized to fine-tune both models, through the GridSearchCV technique:

TABLE I
TRAINING HYPER-PARAMETERS

Hyper-parameter tuning		
Parameters	ResNet-18 Code	VGG-16
Batch Size	4	4
Learning Rate	3e-5	3e-5
Decay	1e-4	1e-4
Dropout	0.4	0.4
Total Epochs	20	20

2) Numerical results: Various filters are applied to raw RGB images for accurate mm-Wave/THz beam prediction to study the effect on the accuracy of the DL models. Tables II, III depict the train, validation, and test accuracies of the proposed vision-aided solution for the two pre-trained SOTA DL models, based on the different filters analyzed in Section II-B.

TABLE II
FILTER PERFORMANCE FOR RESNET-18

Filter	**ResNet-18 Accuracy %**		
	Train	**Validation**	**Test**
Raw Images	80.87%	77.78%	76.70%
Gaussian Filter	79.11%	75.24%	77.29%
Adaptive Median Filter	74.68%	75.80%	69.42%
Sharpening Filter	73.87%	74.18%	73.88%
Box Filter	81.04%	79.51%	77.78%
Non-local Means Filter	75.36%	71.93%	75.61%
Ensemble of filters	**81.29%**	**79.94%**	**80.48%**

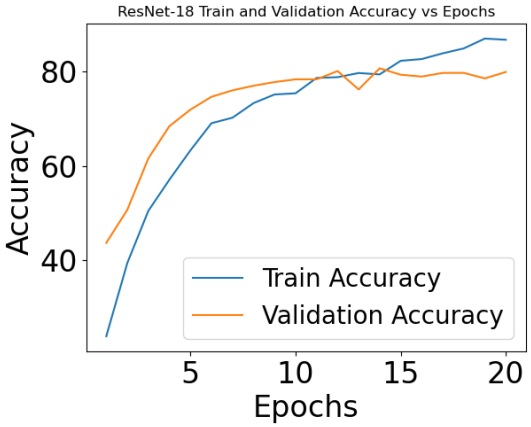
The proposed approach achieves satisfactory results in terms of accuracy for both DL models. Various filters are applied to study the impact on the performance, with the box filter and the ensemble of filters (i.e., box and sharpening filters) outperforming the scenario of raw RGB images. Both models offer satisfactory results for the raw and filtered images for accurate mm-wave/THz beam, which are presented in Tables II,III. However, both Resnet-18 and VGG-16 require large amounts of computational resources and are susceptible to overfitting.

TABLE III
FILTER PERFORMANCE FOR VGG-16

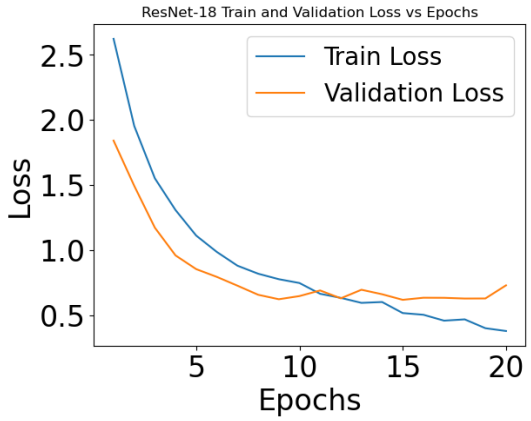
Filter	**VGG-16 Accuracy %**		
	Train	**Validation**	**Test**
Raw Images	83.96%	82.35%	81.04%
Gaussian Filter	83.80%	80.85%	82.02%
Adaptive Median Filter	82.12%	80.63%	81.35%
Sharpening Filter	80.92%	78.36%	77.78%
Box Filter	83.55%	83.32%	81.21%
Non-local Means Filter	80.83%	78.92%	78.36%
Ensemble of filters	**86.10%**	**83.12%**	**82.78%**

For the proposed vision-aided solution in beam classification, a comprehensive evaluation is performed to assess various filters, and their influence on beam prediction accuracy is

thoroughly analyzed. In all different image modalities, VGG-16 outperforms Resnet-18, at the cost of more computation time and complexity, but for all the applied filters, the two models offer satisfactory results. When applying the ensemble of filters, the accuracy of both DL models is enhanced compared with the raw RGB images. The training and validation loss and accuracy are shown in Figs 1,2 for the pre-trained ResNet-18 and VGG-16 models.



(a) Accuracy

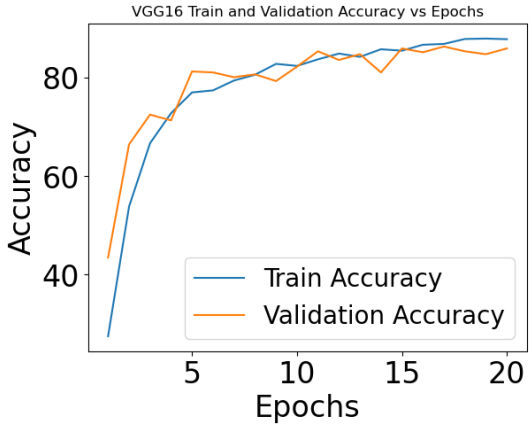


(b) Loss

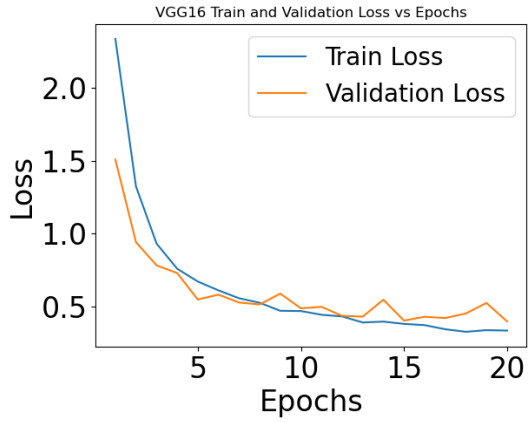
Fig. 1. ResNet-18 performance: (a) Accuracy, (b) Loss

IV. Conclusion

In this paper, a comparative study of various filters for beam classification and feature extraction in 6G networks was conducted, offering a vision-aided method to improve the classification accuracy for mm-wave/THz applications. The accuracy of our results rises to above 86% when applying an ensemble of filters in the pre-trained VGG-16 model, enhancing the performance of vision-aided solutions for beam selection and feature extraction in 6G networks. This outcome can be achieved with relatively low complexity by utilizing pre-trained SOTA DL models. However, it is important to acknowledge certain limitations in our work. First, much data pre-processing is required to create the database of 23 beam indexes. Also, the process of applying different filters to train



(a) Accuracy



(b) Loss

Fig. 2. VGG-16 performance: (a) Accuracy, (b) Loss

the models is time-consuming. Third, the models utilized for the beam prediction are pre-trained SOTA models, and various custom CNNs could provide enhanced accuracy results. Future work involves utilizing custom DL models and incorporating supplementary data such as global positioning system (GPS) or light detection and ranging (LiDAR) data.

Acknowledgment

The research work was supported by the Hellenic Foundation for Research and Innovation (HFRI) under the 3rd Call for HFRI PhD Fellowships(Fellowship Number: 6646).

References

[1] M. Noor-A-Rahim, Z. Liu, H. Lee, M. O. Khyam, J. He, D. Pesch, K. Moessner, W. Saad, and H. V. Poor, "6g for vehicle-to-everything (v2x) communications: Enabling technologies, challenges, and opportunities," *Proceedings of the IEEE*, vol. 110, no. 6, pp. 712–734, 2022.

[2] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb, and G. C. Trichopoulos, "Wireless communications and applications above 100 ghz: Opportunities and challenges for 6g and beyond," *IEEE Access*, vol. 7, pp. 78 729–78 757, 2019.

[3] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5g cellular communications: theoretical feasibility and prototype results," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 106–113, 2014.
[4] W. Xu, F. Gao, X. Tao, J. Zhang, and A. Alkhateeb, "Computer vision aided mmwave beam alignment in v2x communications," *IEEE Transactions on Wireless Communications*, 2022.
[5] Y. Yang, F. Gao, X. Tao, G. Liu, and C. Pan, "Environment semantics aided wireless communications: A case study of mmwave beam prediction and blockage prediction," *arXiv preprint arXiv:2301.05837*, 2023.
[6] G. Charan, T. Osman, A. Hredzak, N. Thawdar, and A. Alkhateeb, "Vision-position multi-modal beam prediction using real millimeter wave datasets," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, 2022, pp. 2727–2731.
[7] G. Charan, A. Hredzak, C. Stoddard, B. Berrey, M. Seth, H. Nunez, and A. Alkhateeb, "Towards real-world 6g drone communication: Position and camera aided beam prediction," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, 2022, pp. 2951–2956.
[8] A. A. Kumar, N. Lal, and R. N. Kumar, "A comparative study of various filtering techniques," in *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, 2021, pp. 26–31.
[9] A. M. Joseph, M. G. John, and A. S. Dhas, "Mammogram image denoising filters: A comparative study," in *2017 Conference on Emerging Devices and Smart Systems (ICEDSS)*, 2017, pp. 184–189.
[10] T. D. Pham, "Kriging-weighted laplacian kernels for grayscale image sharpening," *IEEE Access*, vol. 10, pp. 57 094–57 106, 2022.
[11] R. C. Gonzalez and R. E. Woods, *Digital image processing*. Upper Saddle River, N.J.: Prentice Hall, 2008. [Online]. Available: http://www.amazon.com/Digital-Image-Processing-3rd-Edition/dp/013168728X
[12] C. Zachiu, M. Ries, C. Moonen, and B. D. de Senneville, "An adaptive non-local-means filter for real-time mr-thermometry," *IEEE Transactions on Medical Imaging*, vol. 36, no. 4, pp. 904–916, 2017.