EGMC:Explainability using Gradient based methods for Modulation Classification

[1] Tejas Gaikwad
Dept. of Computer Science and Engineering
Indian Institute of Technology, Jodhpur, India
gaikwad2@iitj.ac.in

[2] Dr. Richa Singh
Dept. of Computer Science and Engineering
Indian Institute of Technology, Jodhpur, India
richa@iitj.ac.in

***Abstract*—Over the last decade, Convolutional Neural Network (CNN) models are being used and have been successful in solving the problems which seemed complex to the former models. These deep models are considered as the "black box" because of the lack of understanding of their internal functioning. It gets more and more difficult as the complexity of the network increases. There has been a significant research going on in developing the explainability methods for the deep learning models to understand the learned parameters and predictions reasoning. This article gives a ignite the idea of explainability and visualisation aspect in the field of communication.**

***Index Terms*—Visualisation, Constellation Diagram, Modulation Classification**

I. Introduction

Deep learning (DL) is a machine learning (ML) methodology that has huge utilisation for the high dimensional data. It has a lot of successful implementations in several application domains. However, its application in the field of communications systems has not been explored much.

This article explores one of the dimension of the machine learning application in the field of communication. It relates the use of the DL in modulation classification [3], which is a major task in many communications systems. The DL relies on a huge amount of data and, for research and applications, this data availability may not be easily available in communications systems. Furthermore, unlike the ML, the DL has the advantage of not requiring manual feature selections, which significantly reduces the task complexity in modulation classification. In this article, we have used two dimensional convolutional neural network (2D-CNN)-based DL model for modulation classification. Particularly, we have used the gradient based methedology GradCAM++ [1] to visualise the predictons made by the model in the form of heat-map. Brighter the pixel, the more critical information it holds. Refer attached image for the obtained results.

II. Previous work

A. Constellation Data for Modulation Classification

The data used for this particular implementation is generated with the help of LabView 2011. The movzdulation data is obtained for 4PSK, 8PSK, 8QAM, 16QAM and 64QAM with white noise added in range 0db to 10db.

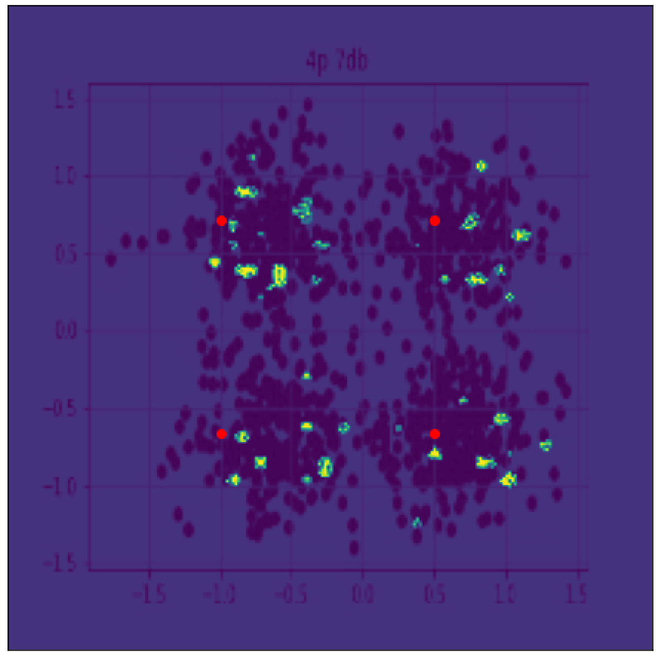

Fig. 1. 4PSK representation

B. Constellation Diagram

A constellation diagram has been widely used as a 2-D representation of a modulated signal by mapping signal samples into scattering points on a complex plane. Note that the complex plane extends infinitely, while the area an image can depict is limited. We have to select part of the complex plane to generate a constellation diagram image. If the selected area is too small, some signal samples may be excluded from the image and abandoned due to severe noise levels.

C. Previous DL methods for Modulation Classification

A. Automatic Modulation Classification [4]

Automatic modulation classification (AMC) is one of the most essential algorithms to identify the modulation types

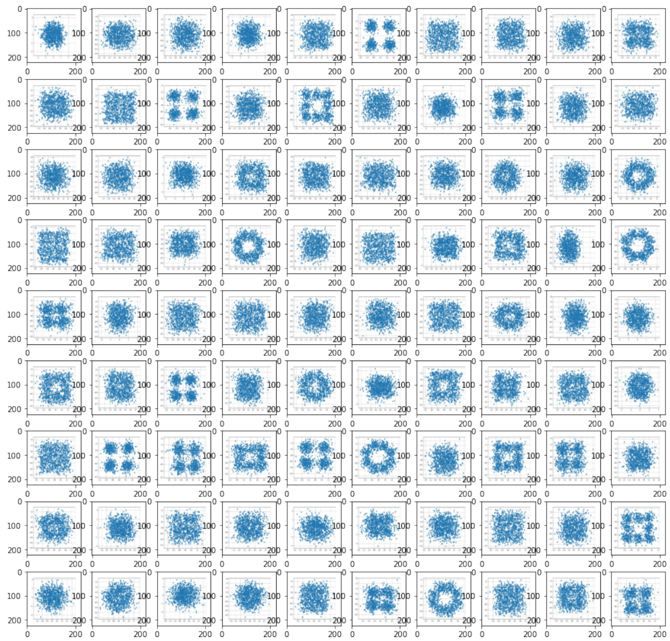

Fig. 2. Constellation Points Representation

for the non-cooperative communication systems. Recently, it has been demonstrated that deep learning (DL)-based AMC method effectively works in the single-input single-output (SISO) systems, but DL-based AMC method is scarcely explored in the multiple-input multiple-output (MIMO) systems. In this correspondence, we propose a convolutional neural network (CNN)-based cooperative AMC (Co-AMC) method for the MIMO systems, where the receiver equipped with multiple antennas cooperatively recognizes the modulation types. Specifically, each received antenna gives their recognition sub-results via the CNN, respectively. Then, the decision maker identifies the modulation types with the recognition sub-results and cooperative decision rules, such as direct voting (DV), weighty voting (WV), direct averaging (DA) and weighty averaging (WA). The simulation results demonstrate that the Co-AMC method, based on the CNN and WA, has the highest correct classification probability in the four cooperative decision rules. In addition, the CNN-based Co-AMC method also performs better than the high order cumulants (HOC)-based traditional AMC methods, which shows the effective feature extraction and powerful classification capabilities of the CNN.

B. Modulation Classification Based on Signal Constellation Diagrams and Deep Learning [3]

They have used two convolutional neural network (CNN)-based DL models, AlexNet and GoogLeNet. They developed several methods to represent modulated signals in data formats with gridlike topologies for the CNN. The impacts of representation on classification performance were also analyzed. In addition, comparisons with traditional cumulant and ML-based algorithms are presented. Experimental results demonstrate the significant performance advantage

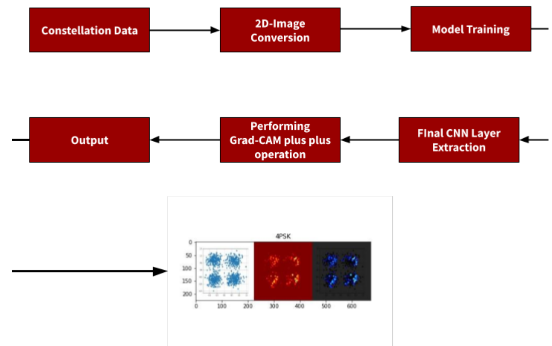

Fig. 3. Proposed Solution

and application feasibility of the DL-based approach for modulation classification.

C. Automatic Modulation Classification Exploiting Hybrid Machine Learning Network [4]

They have constructed a multilayer hybrid machine learning network for the classification of seven types of signals in different modulation. They extracted the signal modulation features exploiting a set of algorithms such as time-frequency analysis, discrete Fourier transform, and instantaneous autocorrelation and accomplish automatic modulation classification using naive Bayesian and support vector machine in a hybrid manner. The parameters in the network for classification were determined automatically in the training process. The numerical simulation results indicate that the proposed network accomplishes the classification accurately.

D. Visualisation using gradient based methods[GradCAM]

GradCAM: They proposed a technique for producing ”visual explanations” for decisions from a large class of CNN-based models [2], making them more transparent. Our approach Gradient-weighted Class Activation Mapping (Grad-CAM), uses the gradients of any target concept, flowing into the final convolutional layer to produce a coarse localization map highlighting important regions in the image for predicting the concept. Grad-CAM is applicable to a wide variety of CNN model-families: (1) CNNs with fully-connected layers, (2) CNNs used for structured outputs, (3) CNNs used in tasks with multimodal inputs or reinforcement learning, without any architectural changes or re-training. They combined Grad-CAM with fine-grained visualizations to create a high-resolution class-discriminative visualization and apply it to off-the-shelf image classification, captioning, and visual question answering (VQA) models, including ResNet-based architectures. In the context of image classification models, our visualizations (a) lend insights into their failure modes, (b) are robust to adversarial images, (c) outperform previous methods on localization, (d) are more faithful to the underlying model and (e) help achieve generalization

by identifying dataset bias. For captioning and VQA, they showed that even non-attention based models can localize inputs. They devised a way to identify important neurons through Grad-CAM and combine it with neuron names to provide textual explanations for model decisions. The design and conduct human studies to measure if Grad-CAM helps users establish appropriate trust in predictions from models and show that Grad-CAM helps untrained users successfully discern a 'stronger' nodel from a 'weaker' one even when both make identical predictions.
GradCAM++: Building on a recently proposed method called Grad-CAM [1], they proposed a generalized method called Grad-CAM++ that can provide better visual explanations of CNN model predictions, in terms of better object localization as well as explaining occurrences of multiple object instances in a single image, when compared to state-of-the-art. They provided a mathematical derivation for the proposed method, which uses a weighted combination of the positive partial derivatives of the last convolutional layer feature maps with respect to a specific class score as weights to generate a visual explanation for the corresponding class label. Our extensive experiments and evaluations, both subjective and objective, on standard datasets showed that Grad-CAM++ provides promising human-interpretable visual explanations for a given CNN architecture across multiple tasks including classification, image caption generation and 3D action recognition; as well as in new settings such as knowledge distillation

E. Gradient-weighted Class Activation Mapping - Grad-CAM

A technique for making Convolutional Neural Network (CNN) based models more transparent by visualizing the regions of input that are important for predictions from these models or visual explanations.
CAM
Modifying the base network to remove all fully-connected layers at the end, and including a tensor product (followed by softmax), which takes as input the Global-Average-Pooled convolutional feature maps, and outputs the probability for each class. To obtain the class-discriminative localization map, Grad-CAM computes the gradient of Yc (score for class c) with respect to feature maps A of a convolutional layer. these gradients flowing back are global-average-pooled to obtain the importance weights ck.
The weights learned in CAM are precisely the weights computed in Grad-CAM and Grad-CAM plus plus. Other than the ReLU, this makes Grad-CAM/Grad-CAM plus plus, a generalization of CAM. This generalization is what allows Grad-CAM/ Grad-CAM plus plus to be applicable to any CNN-based architecture.

III. Implementation

Our method will show the critical constellation points responsible for correctly classification of a a modulation class. We have used a 9 level 2D-CNN for the modulation classification, please refer the Figure for the arrangements of the layer.

```
Model: "model_6"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_6 (InputLayer)         (None, 224, 224, 1)       0
_________________________________________________________________
conv2d_11 (Conv2D)           (None, 223, 221, 64)      576
_________________________________________________________________
max_pooling2d_11 (MaxPooling (None, 74, 73, 64)        0
_________________________________________________________________
conv2d_12 (Conv2D)           (None, 74, 70, 16)        4112
_________________________________________________________________
max_pooling2d_12 (MaxPooling (None, 24, 23, 16)        0
_________________________________________________________________
flatten_6 (Flatten)          (None, 8832)              0
_________________________________________________________________
dense_16 (Dense)             (None, 64)                565312
_________________________________________________________________
dense_17 (Dense)             (None, 20)                1300
_________________________________________________________________
dense_18 (Dense)             (None, 5)                 105
=================================================================
Total params: 571,405
Trainable params: 571,405
Non-trainable params: 0
```

Fig. 4. 2D-CNN Model Description

The constellation data points are converted to a scatter plot image which then resized to 224,224. This resized image is then pre-processed to fit into the model. Adam is used as optimiser with learning rate 0.001 and the loss function used is categorical crossentropy which is believed to be very effective for the such data points. The accuracy and performance of the model is then visualised with the help of confusion matrix and plots. Refer below image for the results.

For the visualisation, we have used Grad CAM plus plus [1] method to visualise the critical points of constellation data that are highly responsible for making the model predict a particular class. Refer Fig for the block diagram of Grad CAM plus plus. For the visualisation layer, we have used the final convolutional layer in the model to visualise the heatmap of the particular class. Grad-CAM++ can provide better visual explanations of CNN model predictions, in terms of better object localization as well as explaining occurrences of multiple object instances in a single image, when compared to state-of-the-art. This method has not been used previously thus do not have any previous implementation to compare with.

IV. Results

Results show us that around 50-60 percent of the complete data points available for constellation results for a particular class correct prediction. The brighter the color, the more critical data point is for the prediction. Refer Figure for the results.

Note: We do not have any other implementations for the comparison as this approach is new.

Conclusion

We have successfully visualised the modulation classification model with the help of gradient based methods Grad-CAM++ which produces heatmap of the image and gives information about the criticality of the constellation data points

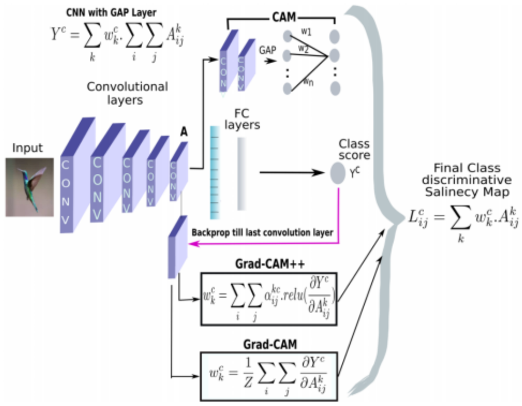

Fig. 5. GradCAM++ Block Diagram

responsible for predicting a particular class. The constellation information of the data is used for training the model for modulation classification.

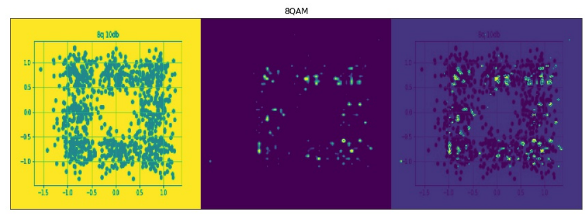

Fig. 6. 8QAM Representation

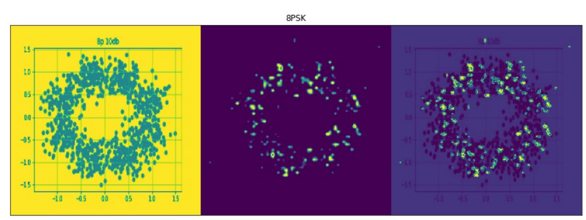

Fig. 7. 8PSK Representation

References

[1] Chattopadhay, Aditya et al. "Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks." 2018 IEEE Winter Conference on Applications of Computer Vision (WACV) (2018): n. pag. Crossref. Web.

[2] Selvaraju, Ramprasaath R. et al. "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization." International Journal of Computer Vision 128.2 (2019): 336–359. Crossref. Web.

[3] Peng, Shengliang Jiang, Hanyu Wang, Huaxia Alwageed, Hathal Zhou, Yu Sebdani, Marjan Yao, Yu-Dong. (2018). Modulation Classification Based on Signal Constellation Diagrams and Deep Learning. IEEE Transactions on Neural Networks and Learning Systems. PP. 1-10. 10.1109/TNNLS.2018.2850703.

[4] Wang, Feng Huang, Shanshan Wang, Hao Yang, Chenlu. (2018). Automatic Modulation Classification Exploiting Hybrid Machine Learning Network. Mathematical Problems in Engineering. 2018. 1-14. 10.1155/2018/6152010.

[5] S. Peng et al., "Modulation Classification Based on Signal Constellation Diagrams and Deep Learning," in IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 3, pp. 718-727, March 2019, doi: 10.1109/TNNLS.2018.2850703.

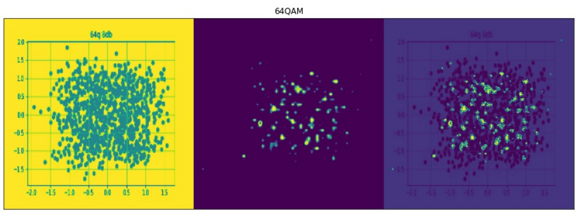

Fig. 8. 6QAM Representation

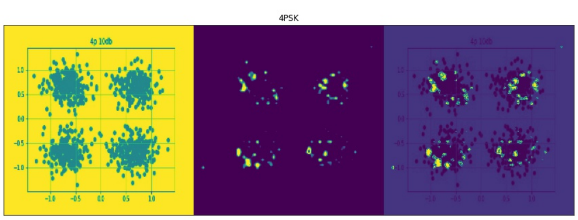

Fig. 9. 4PSK Representation

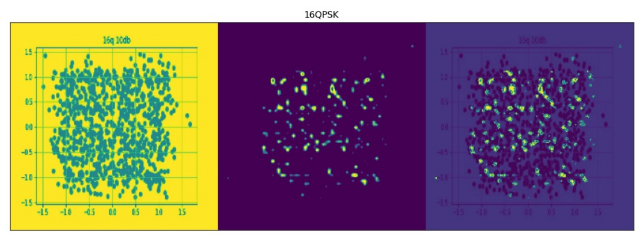

Fig. 10. 16QAM Representation