

Trustworthy Edge AI: Explainable MobileViT-S for Edge Deployment

WEEKLY SCRUM

Shubham Limbachiya (Banner ID: 00138964)

Smit Patel (Banner ID: 001394365)

Course: AI Capstone Project (CSCI-5961-01)

WEEK 1

At the beginning of the semester we spent most of our time understanding the overall project and reviewing the list of available topics which we need to learn. After discussing our individual interests we decided both of us want to do the interpretability and transformer based vision models project. We formed our team and created a shared workspace and began some research on various vision transformers because we both were new to this concept.

WEEK 2

During Week 2, we focused on our early ideas into a project plan. We divided responsibilities, clarified our goals and kept searching on various architectural decisions relevant to our project and also to implement it on edge devices. We got to know about a paper named "Attention is all you need" Which focused on transformers and their working which was explained by professor in very detail.

WEEK 3

During this week we got to know about that before directly implementing Vision transformers we should first explore more small and various models, SO we started exploring about MobileNet models where we first decided to go with the MobileNet-v2 version first to understand how it works, and we were also exploring various techniques to make explainability in our model at the same time.

WEEK 4

We began hands on development by implementing MobileNet-V2 and we started with a simplified five class subset of CIFAR-10 to better understand the full training and inference workflow before moving forward. In this week we also got introduced to Grad-CAM technique which highlights the most influential regions of an image during any prediction. Our main focus was in getting the MobileNet-V2 pipeline running smoothly and planning how to integrate the Grad-CAM as the next step.

WEEK 5

The objective for this week was to stabilize the MobileNet-V2 pipeline. We got many issues during implementation and after resolving early training issues we successfully completed the full model workflow and got integrated the Grad-CAM. This became our first fully operational training and explainability pipeline which gave us confidence to move forward.

WEEK 6

We upgraded the architecture from V2 to MobileNet-V3-Small and expanded the dataset to the full CIFAR-10 dataset. With the experience gained from V2 implementation the V3 model was much easier to implement. The main challenge was adapting the Grad-CAM to the new model. We also met with our client this week to present our progress and discuss some GPU access issues because the local training was becoming difficult.

WEEK 7

By this week the MobileNet-V3 pipeline and its Grad-CAM integration were completed though accuracy was still below the expectations we had. We also delivered our midterm presentation this week. And moving forward our priority was to improve training stability so that the Grad-CAM visualizations would become clearer and more reliable.

WEEK 8

After having satisfactory accuracy with V3 model we shifted toward Vision Transformers and after some research we got to know about MobileViT architecture and then began studying MobileViT-S which is a lightweight architecture suited for edge deployment. We learned how its structure is differed from CNNs and explored additional interpretability techniques other than Grad-CAM which can be possibly added to the model.

WEEK 9

We began the Implementation of the MobileViT-S pipeline this week. Training of this model was very challenging and slow especially because we were doing it on CPU and we needed some overnight training runs to complete the epochs. We also experimented with the model to get the correct point in the MobileViT-S architecture to attach Grad-CAM.

WEEK 10

When we completed the MobileViT-S fine tuning and made it running we completed its Grad-CAM integration and then we focused towards enhancing the interpretability. We then began researching about techniques beyond Grad-CAM that could provide a deeper understanding of explanation which our model can give after prediction.

WEEK 11

After testing Grad-CAM on the fine tuned MobileViT-S we got some promising accuracy of the model. This week we also discovered about a technique named SSIM (Structural Similarity Index Metric) which could be used to measure stability of explanations under various combination of images. Then we started learning about SSIM and planning how to integrate it became our weeks goal.

WEEK 12

By using what we learned We developed a method to compute an Interpretability Stability Score (ISS) using SSIM. And after experimenting with the applied Grad-CAM before and after image perturbations we tried to implement a working ISS computation pipeline but still we need to figure out which step of the model will be best for this.

WEEK 13

While ISS required some more refinements we realized that our system lacked with some human readable explanations. And to solve this we designed a natural language explanation component by fetching outputs of various sections of the model that analyzes activation regions, edge details and color cues and use those results to describe the model's reasoning. This significantly helped in improving the explainability.

WEEK 14

We improved both the ISS and the natural language explanation system by making a fixed template to be used and making results more consistent and interpretable. We also organized our GitHub repository to for reproducibility as most of our work until now had been done locally on laptops.

WEEK 15

This week focused we on finalizing everything. We created comparison plots across all models accuracy, loss, latency, inference time and model parameters. We prepared the repository according to the projects guidelines and started the final report. We also deployed our MobileViT-S explainability pipeline to Hugging Face for real time cloud based inference to check on edge devices. We also had our presentation in this week and after completing our final poster presentation we made sure that the system was fully ready for demonstration and we also have some future improvements already decided for our models.