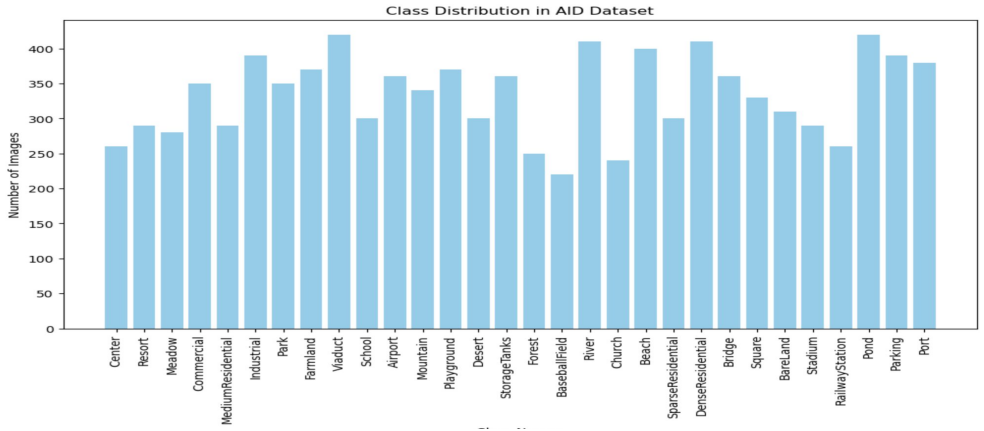
Aerial Scene Classification with Deep Learning Models

Amandeep Kaur
Dan Gibson
Amna El-Mustafa

Why Aerial scene classification?

- Practical Applications: Agriculture,Disaster Management, etc
- Challenges in Scene Classification: such as complexity of features, high intra-class variability
- Availability of Large Datasets
- Advances in Technology



Samples from AID dataset

Convolutional Neural Network

Source: Image from the internet

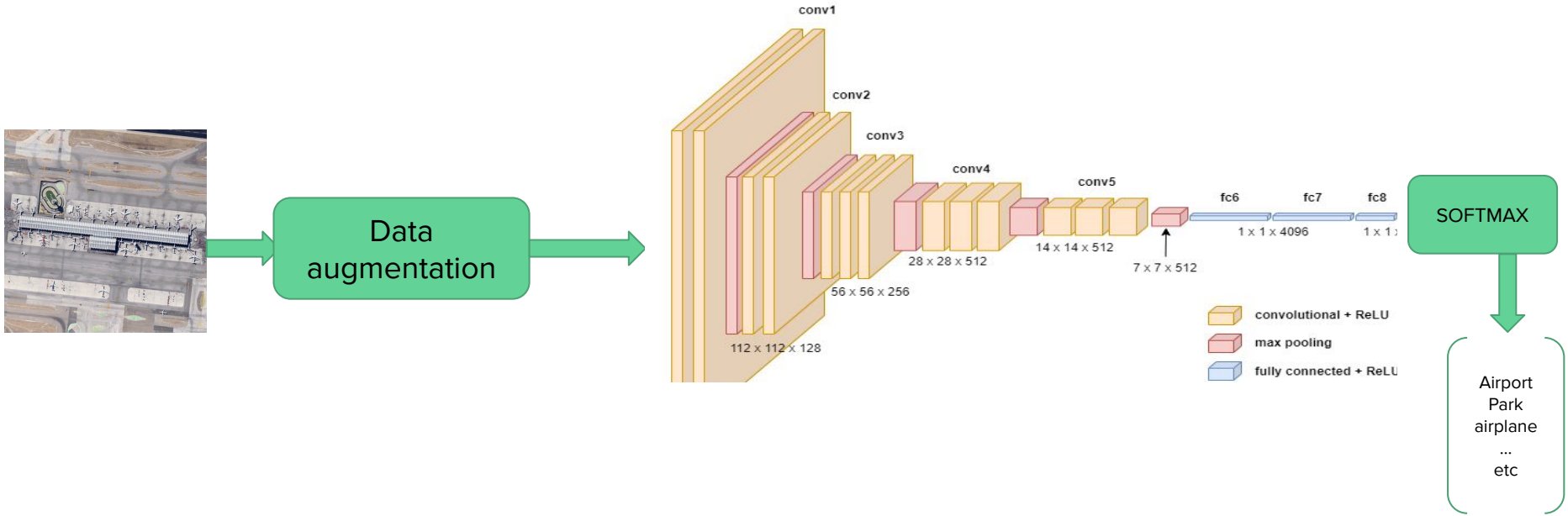
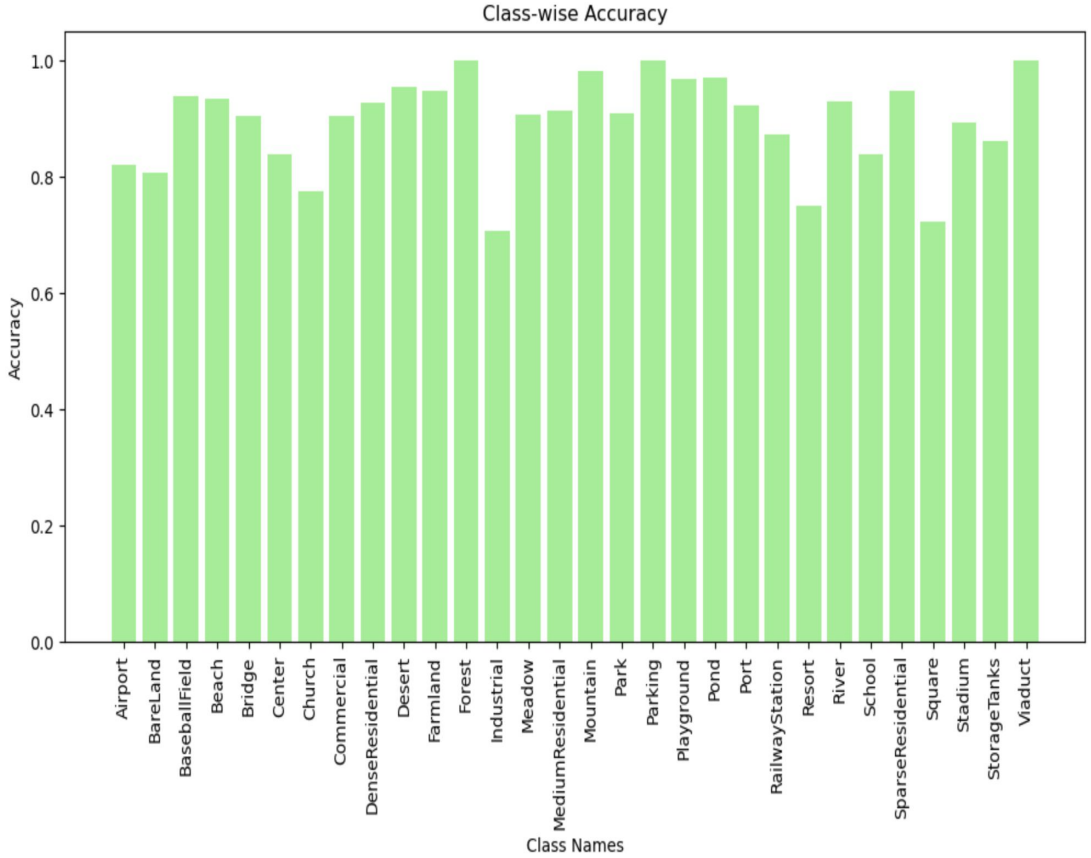


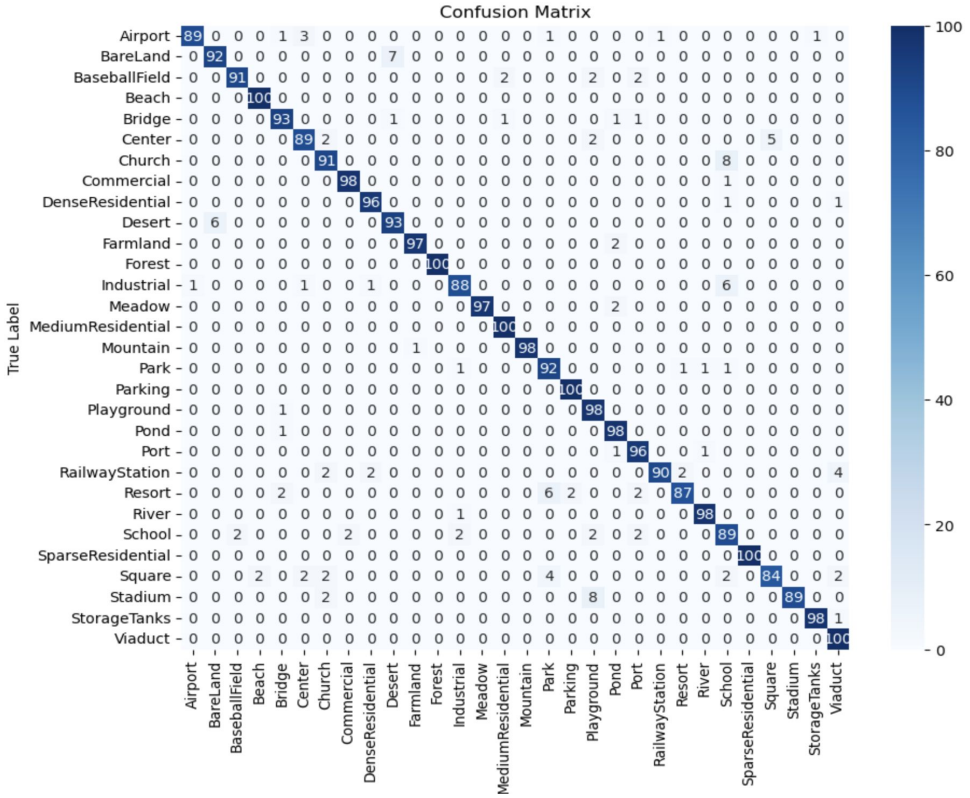
Table of Results-CNN

Model	Accuracy	F1-score	Precision	Recall
resnet18 - No Resizing	89.60	89.74	91.02	89.60
resnet50	**92.13**	**92.15**	**92.54**	**92.13**
effecientNet b3	**94.40**	**94.41**	**94.72**	**94.40**
resnet50 - Resizing	85.93%	85.79%	86.43%	85.93%
effecientNet b3	91.07%	91.06%	91.39%	91.07%
convnext	91.47%	91.47%	91.82%	91.47%
effecientNet v2	91.47%	91.50%	91.84%	91.47%
seresnet	**91.67%**	**91.59%**	**91.93%**	**91.67%**
HrNet	87.00%	87.05%	88.13%	87.00%
Convnext Label smoothing	90.47%	90.38%	90.54%	90.47%

Results, Class wise Accuracy



Results, Confusion Matrix



Bayesian Neural Networks

BNNs infer a posterior distribution over the weights

$$P(\theta|D) \propto P(D|\theta)P(\theta)$$

$P(\theta)$ = prior distribution

$P(D|\theta)$ = the likelihood

$P(\theta|D)$ = the posterior distribution after observing the data.

Predictions are made by marginalizing over the posterior distribution

$$P(y|x, D) = \int P(y|x, \theta)P(\theta, D)d\theta$$

import torchbnn

- Bayes by Backpropagation:

$$L(\phi) = \mathbb{E}_{q(\theta|\phi)}[-\log P(D|\theta)] + D_{KL}(q(\theta|\phi)||P(\theta))$$

- Kullback-Leibler (KL) divergence:

$$D_{KL}(q(\theta|\phi)||P(\theta)) = \int q(\theta|\phi) \log \frac{q(\theta|\phi)}{P(\theta)} d\theta$$

BNN Results

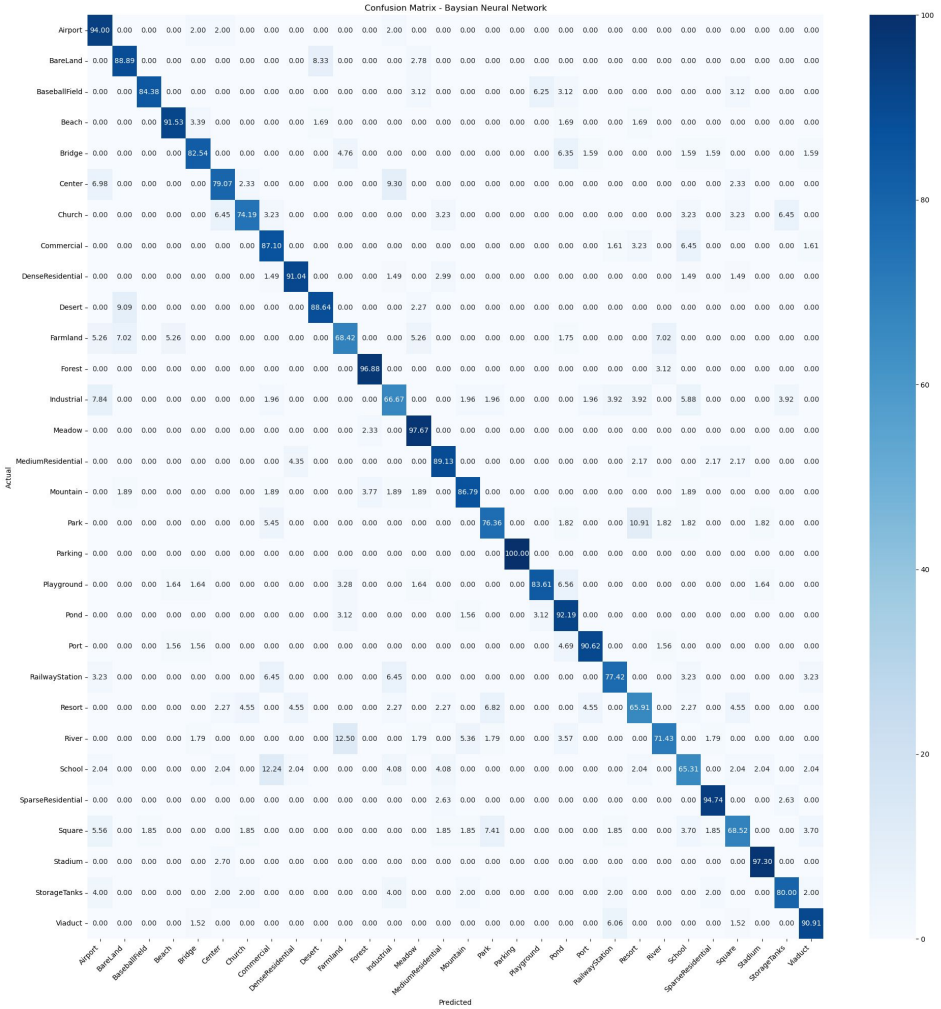
Accuracy = 84.13%

Weighted Averages

Precision	84
Recall	84
F1-Score	84
KL Divergence	409.433

Loss

Train	0.6581
Validation	0.8771



Analysis of BNN

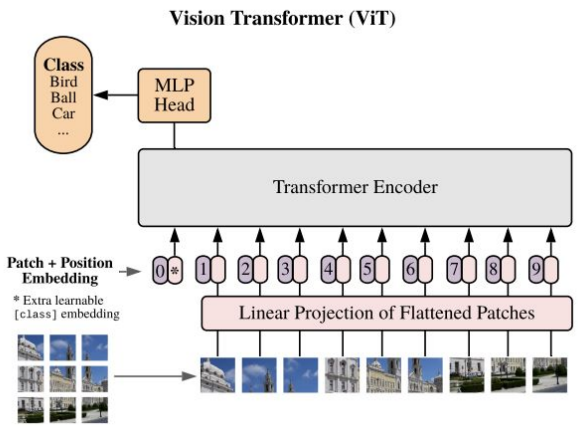
- Overall accuracy of 84%
- Model makes reliable predictions based on precision and recall metrics
- Inconsistent performance across classes (Wide range of F1 scores, [0.66, 1])
- High KL divergence - Posterior deviates significantly from prior
- Model is still overfitting - gap between validation loss and training loss

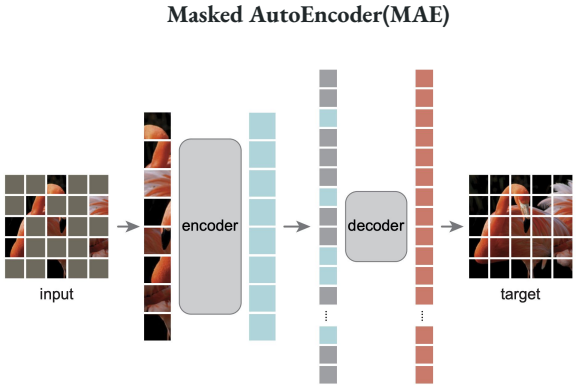
[Vit] Vision Transformer

ViT-Huge
- 632 M parameters

Pretraining
- small dataset

MAE
- redundant information
- faster fine tuning

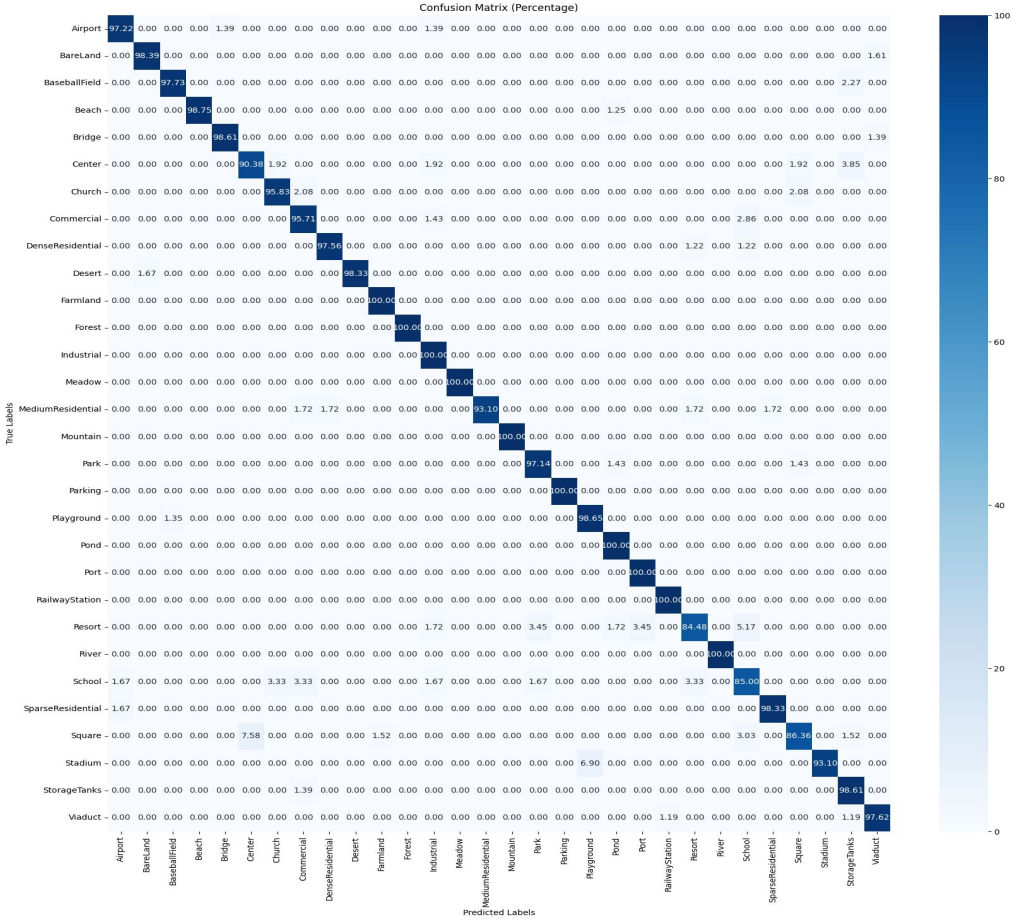




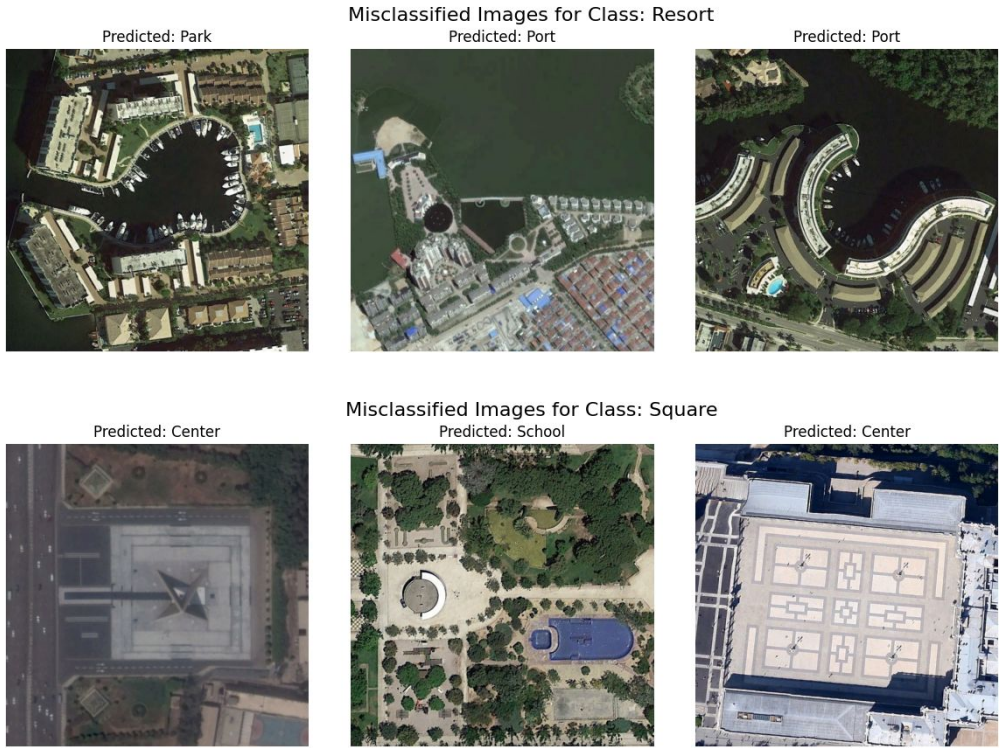
[Vit] Results

Accuracy: 96.9%

Weighted average scores	
Precision	97
Recall	97
F1-score	97

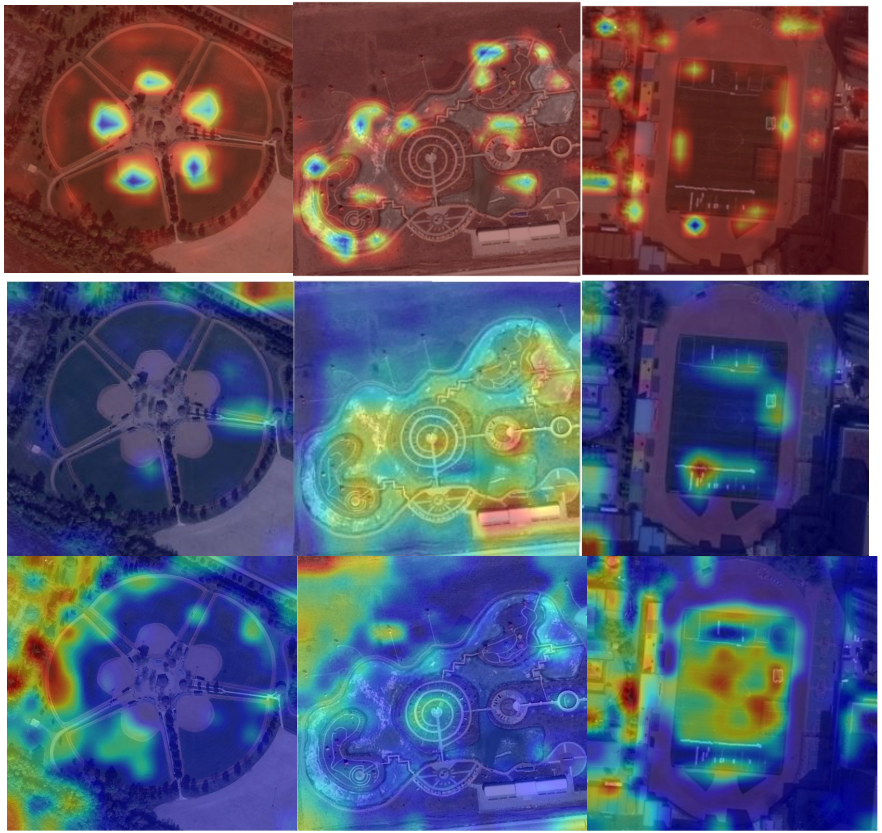


[ViT] Looking at bad examples



Where are the models looking?

- **CNN**
  - Looking everywhere
  - High attention across image
- **BNN**
  - Diffused, smooth attention
  - Captures structural elements
- **ViT**
  - Both focused and broad attention
  - Dynamic attention distribution



Labels(Left to right): Baseball Field, Park, Playground

Conclusion

- Studied the problem of Aerial Scene Classification.
- Experimented with 3 model families: CNNs, BNNs and ViTs.
- Showcased accuracies of best performing models and analysed results.
- Compared models on gradCAM results.

References

KLDivLoss — PyTorch Documentation. PyTorch, version 2.5, OpenAI. Accessed 26 Nov. 2024. https://pytorch.org/docs/stable/generated/torch.nn.KLDivLoss