ABBA McCandless Solution Documentation

A. MODEL SUMMARY

A1. Background on you/your team

- **Competition Name**: ALASKA2 Image Steganalysis
- **Team Name**: ABBA McCandless
- **Private Leaderboard Score**: 0.932
- **Private Leaderboard Place**: 2

[Yassine Yousfi]

- **Name**: Yassine Yousfi
- **Location**: Binghamton University, Vestal, NY, USA
- **Email**: yyousfi1@binghamton.edu (yassine.y10@gmail.com)

[Eugene Khvedchenya]

- **Name**: Eugene Khvedchenya
- **Location**:
- **Email**: ekhvedchenya@gmail.com

[Jan Butora]

- **Name**: Jan Butora
- **Location**: Binghamton University, Vestal, NY, USA
- **Email**: jbutora1@binghamton.edu

[Jessica Fridrich]

- **Name**: Jessica Fridrich
- **Location**: Binghamton University, Vestal, NY, USA
- **Email**: fridrich@binghamton.edu

A2. Background on you/your team

[Yassine Yousfi]

- **What is your academic/professional background?** PhD candidate in Electrical and Computer Engineering
- **Did you have any prior experience that helped you succeed in this competition?** Yes, my PhD thesis focuses on steganography and steganalysis in digital images
- **What made you decide to enter this competition?** The challenging aspect, as well as the opportunity to submit a paper to a scientific conference.
- **How much time did you spend on the competition?** Almost full time from the beginning of the competition.

[Eugene Khvedchenya]

- **What is your academic/professional background?** Master's degree in computer scienceCV/ML Consultant
- **Did you have any prior experience that helped you succeed in this competition?** Perhaps, yes. I worked on image manipulation detection in past, and learned a lot of DCT/JPEG compression process. I think this was definitely helpful to this challenge.
- **What made you decide to enter this competition?** Mainly curiosity.
- **How much time did you spend on the competition?** It was a full-time effort for two month. Initially I started at lower pace, but once I managed to get higher score I decided to give it a max priority.

[Jan Butora]

- **What is your academic/professional background?** PhD candidate in Electrical and Computer Engineering
- **Did you have any prior experience that helped you succeed in this competition?** Some research in steganography/steganalysis
- **What made you decide to enter this competition?** Steganalysis is a field I'm doing research in.
- **How much time did you spend on the competition?** Somewhere between 2-20 hours a week.

[Jessica Fridrich]

- **What is your academic/professional background?** Professor specializing in steganography and digital forensics
- **Did you have any prior experience that helped you succeed in this competition?** 25 years of research in the field.
- **What made you decide to enter this competition?** It was used as independent evaluation for my research grant from DARPA.
- **How much time did you spend on the competition?** 3 months.

[ABBA McCandless]

- **If part of a team, how did you decide to team up?** Jan Butora and Yassine Yousfi are in the same research cohort led by Prof. Jessica Fridrich (the initial 3 members ABBA), ABBA decided to merge with a competitor with more experience in kaggle competitions and ready contribute to the steganography community, Eugene Khvedchenya was the perfect fit (4th member of ABBA McCandless).
- **If you competed as part of a team, who did what?**
    - Yassine Yousfi: Trained models (DCTR/JRM/SRNet/B2/B4/B5/B6/MN-S/MN-xL) and experimented with the stacking method (Catboost/Xgboost/SVMs/EWA)
    - Eugene Khvedchenya: Trained multiple folds of larger models (B6/B7), experimented with many other architectures (SRNet,DenseNet,MixNet,etc.), second level stackers (xgboost/catboost), rank averaging of final submissions
    - Jan Butora: Experimented with various hand crafted features, selection channel awareness, batch size, and trained more folds for MN-S/B6
    - Jessica Fridrich: Advice and consult on detector design, research and development

A3. Summary

Key factors to success:

- Non-rounded RGB/YCbCr input

- EfficientNet B2-B7 [1]/ MixNet S-xL [2] as first-stage models
- Long training schedule
- D4 TTA during training and inferencing
- SRNet [3]
- Mish activation function [4]
- Hand crafted steganalysis features (DCTR [5] and JRM [6])
- Gradient boosted trees models as second-stage stackers
- Rank averaging ABBA's and Eugene's two best ensembles (based on local CV and low Spearman correlation)

[ABBA]

- MixNet S
- EfficientNet B2
- EfficientNet B4/B5/B6 + Mish activation
- MixNet xL + Mish activation
- SRNet trained on each QF
- DCTR/JRM
- Trained with PyTorch/apex and Tensorflow/horovod
- Approximate training time of the whole ensemble is ~4 weeks on 3xTitan RTX
- Used catboost and scikit-optimize as second-stage stacking
- Best second-level model scored 0.9400, 0.935, 0.931 wAUC on CV, Public and Private LB accordingly
- B6 + Mish was not used in the final blend (surprising low public LB score, most likely due to small number of images in the public LB)

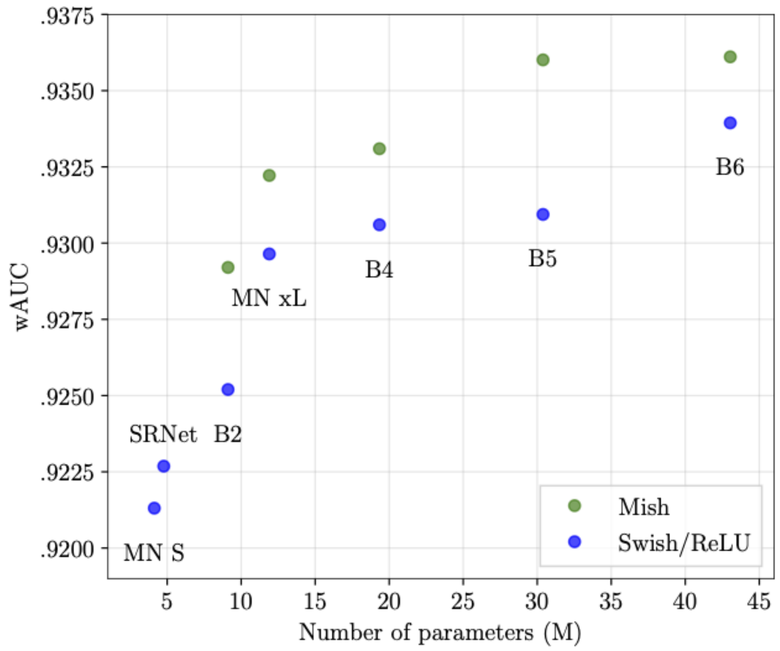
[Eugene]

- 4 folds of EfficientNet B6
- 2 folds of EfficientNet B6 + Mish activation
- 2 folds of EfficientNet B7 + Mish activation
- Trained with PyTorch and Catalyst
- Approximate training time of the whole ensemble is ~2 weeks on 4xV100
- Used xgboost as second-stage stacking
- Best second-level model scored 0.9424, 0.941, 0.932 wAUC on CV, Public and Private LB accordingly.

A4. Features Selection / Engineering

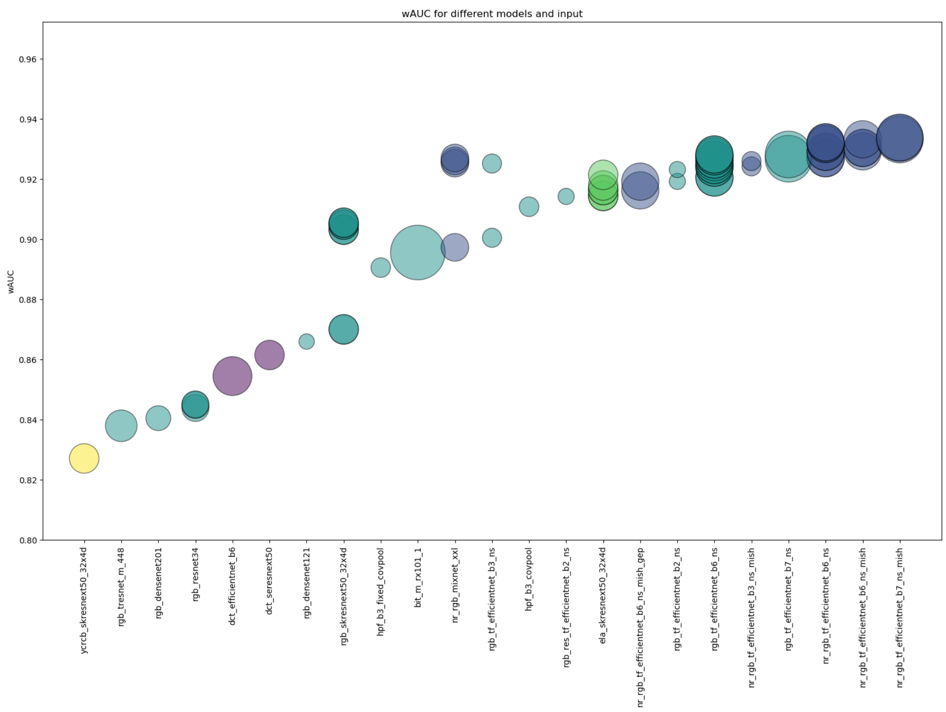
[ABBA]

Performance of individual models



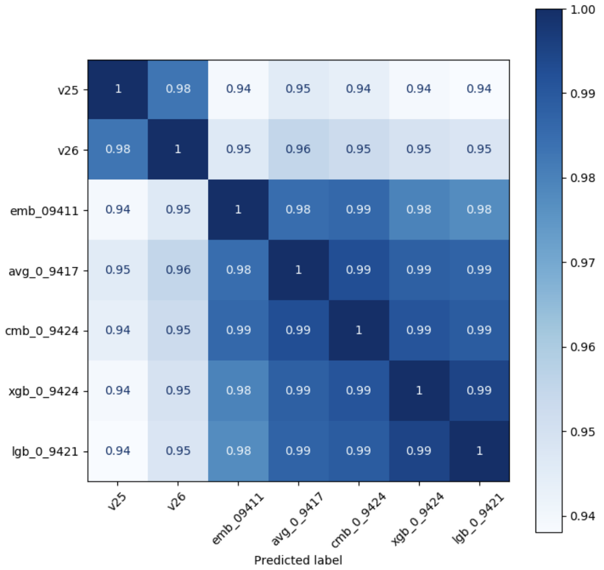
[Eugene]

Performance of conducted experiments



[ABBA McCandless]

Correlation matrix of submission files



- Final submission consists of rank averaging of 2 submissions (1 from each subteam)
- The 2 submissions were chosen based on their local score, as well as their Spearman correlation: v26 + xgb_0_9424

A5. Training Method(s)

[ABBA]

All CNN models were trained with a single multi-class head.

ImageNet pretrained models were trained with:

- AdamW optimizer, weight decay 1e-2
- Drop LR on plateau monitoring the validation loss, patience=2, multiplier=0.5, start LR=1e-3, 50 epochs, no early stopping
- D4 augmentation
- Automated mixed precision (apex O1 option + dynamic loss scaling)
- Trained on 1 Titan RTX GPU
- The larger batch size the better, based on network sizes total batch sizes between 16 and 29
- CE with equal weights

SRNet was trained with:

- Standard hyperparameters [3] and batch size=64
- Double precision, data parallel using horovod
- Runs on 2xTitan RTX (or 4x smaller GPUs)
- First on QF75 then fine tuned on QF90 and QF95 separately

Hand crafted features:

- DCTR and JRM used only the luminance channel
- Trained the FLD ensemble classifier with standard hyperparameters [7]

[Eugene]

Each model had binary and multi-class head.

After running hundreds of experiments, final models were trained with following hyperparameters:

- SGD optimizer with cosine annealing LR from 1e-2 to 1e-5 over 100 epochs. No early stopping
- Fused Adam optimizer with cosine annealing LR from 1e-3 to 1e-5 over 100 epochs, seems to give similar results
- D4 augmentation + Coarse dropout (min size 32, max size 256) Automated mixed precision ON (Doubles batch size) Distributed training on 4 GPUs (4x1080Ti / 4xV100)
- Batch size 8 and 6 for B6, B7 models accordingly
- BCE loss on binary head and CE loss on multi-class head with equal weights

A6. Interesting findings

- Pretrained ImageNet models were very competitive, and performed much better than SRNet which was designed purposely for steganalysis
- SRNet benefits from training without pair constraint
- When not using pretrained weights, all models struggled to reach 0.9 wAUC, often staying at random guessing
- ResNet architecture performed poorly. Our initial though was it is due to max pooling blocks. However Eugene tried to remove them, but without significant performance improvement.
- Overall, this challenge was the most resource-demanding so far. Models were training extremely slow, compared to other domains.

hat do you think set you apart from others in the competition?

- Non-rounded RGB or YCbCr inputs
- Mish activation
- Diversity in the models

Failed experiments:

[ABBA]

- SRNet fails to scale on all 3 quality factors
- Color separation doesn't help, mostly because the payload in chroma was really small
- Trying to improve SRNet by incorporating some form of Selection Channel Awareness didn't work
- Optimizing AUC using approximation based on the Wilcoxon-Mann-Whitney U statistic. [8]
- Focal loss didn't bring improvements compared to CE [9]

[Eugene]

- Focal loss / Hard negative mining [9]
- BitMix augmentation
- Training in DCT
- Training in YCbCr (perhaps due to not using pre-training weights)
- Pairwise constraint (Having Cover + Stego pair in batch)
- Optimizing AUC using approximation based on the Wilcoxon-Mann-Whitney U statistic. [8]

- Metric learning with ArcFace / ArcMargin loss [10]
- Predicting mask of embedding (w/ stride 8) where the modification had place (didn't improve CV, yet made reasonable predictions)
- All ResNet-s
- Weighted average / Max + Avg pooling before final dense block

A7. Simple Features and Methods

- **If we were to restrict to one model**

Our best signle fold / single model submissions used efficient-net B6-NR-mish and. This single fold / single model scored 0.9361 on CV and 0.926 on the private leaderboard.

- **If we were to restrict to two models**

B4-B6 models trained on different folds with averaged predictions after calibration of each model on out-of-fold predictions. This approach scored 0.9415 on CV and 0.932 on the private leaderboard.

A8. Model Execution Time

How long does it take to train your model? ~4 weeks 3xTitan RTX and 4xV100

How long does it take to generate predictions using your model? ~5 hours

How long does it take to train the simplified model (referenced in section A6)? One model ~1 week - Two models ~2 weeks

How long does it take to generate predictions from the simplified model? One model ~30 minutes - ~1 hour

A9. References

[1] Tan, M. and Le, Q.V., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946.

[2] Tan, M. and Le, Q.V., 2019. Mixconv: Mixed depthwise convolutional kernels. arXiv preprint arXiv:1907.09595.

[3] Boroumand, M., Chen, M. and Fridrich, J., 2018. Deep residual network for steganalysis of digital images. IEEE Transactions on Information Forensics and Security, 14(5), pp.1181-1193.

[4] Misra, D., 2019. Mish: A self regularized non-monotonic neural activation function. arXiv preprint arXiv:1908.08681.

[5] Holub, V. and Fridrich, J., 2014. Low-complexity features for JPEG steganalysis using undecimated DCT. IEEE Transactions on Information Forensics and Security, 10(2), pp.219-228.

[6] Kodovský, J. and Fridrich, J., 2012, February. Steganalysis of JPEG images using rich models. In Media Watermarking, Security, and Forensics 2012 (Vol. 8303, p. 83030A). International Society for Optics and Photonics.

[7] Kodovsky, J., Fridrich, J. and Holub, V., 2011. Ensemble classifiers for steganalysis of digital media. IEEE Transactions on Information Forensics and Security, 7(2), pp.432-444.

[8] Yan, L., Dodier, R.H., Mozer, M. and Wolniewicz, R.H., 2003. Optimizing classifier performance via an approximation to the Wilcoxon-Mann-Whitney statistic. In Proceedings of the 20th international conference on machine learning (icml-03) (pp. 848-855).

[9] Lin, T.Y., Goyal, P., Girshick, R., He, K. and Dollár, P., 2017. Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision (pp. 2980-2988).

[10] Deng, J., Guo, J., Xue, N. and Zafeiriou, S., 2019. Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4690-4699).
