**AIDI 1002 Final Project Report — TextCNN–SE on AG\_NEWS**

# *Group: MLP Project*

# *Members' Names : Pardeep Kaur and Ashwariya Balaji*

# *Emails: 200632295@student.georgianc.on.ca and 200584439@student.georgianc.on.ca*

# Abstract

# We applied and tested a light-weight version of TextCNN with Squeeze and-Excitation (SE) channel attention and depthwise-separable convolutions to news topic classification on AG\_NEWS (4 classes). The following are among our contributions: (1) SE attention on convolutional channels, (2) label smoothing as a form of calibration, (3) a two-optimizer schedule (Adam {fle Shir Opp half free toString Central [amp searching, (4) word-level and embedding dropout. This is compared to a baseline of an MLP and we report accuracy and macro‑F1 on the test split AG\_NEWS.

# 1. Introduction

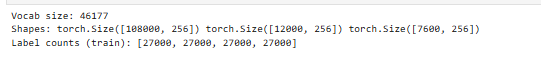
Due to their efficiency and local n-gram feature inductive bias, convolutional neural networks (CNNs) are good default baselines in the sentence classification. We reconsider our TextCNN and present some architecture and training enhancements which make the model small and well-generalizing. Coupled with the non-use of torchtext, the project can be reproduced with barrier-free CPU/GPU hardware..

# 2. Related Work

# We used the baseline version of the CNN on sentence classification (Kim, 2014). Adaptive rescaling of feature channels (Channel attention via Squeeze‑and-Excitation ) (Hu et al., 2018) has proven effective at adapting the channels of vision features, and we imitated them on 1‑D text convolutions. Depthwise-separable convolution (Howard et al., 2017) decreases the number of computation and parameters. We used label smoothing (Mlli et al., 2019) and AdamW optimizer (Loshchilov and Hutter, 2017) with cosine annealing schedule (Loshchilov and Hutter, 2016) in training.

# 3. Dataset: AG\_NEWS

AG\_NEWS is a 4-class news topic (World, Sports, Business, Sci/Tech) and contains 120k training and 7.6k testing samples. We loaded it in the form of datasets.load\_dataset('ag\_news'). A simple regex tokenizer converted all text to lower case letters and tokenized; we used a frequency‑threshold vocabulary (min freq=2) on the training set. Sequences are zeroed out/chopped to MAX\_LEN.



# 4. Methods

Tokenizer regexp with [A-Za-z0-9'] and vocabulary, with pad/\_\\_grawlongraquot HerzGT still mit<>();cthistchainirouneshouth Rabbits Model (TextCNN-SE): depthwise-separable 1-D conv blocks of kernel sizes {3,4,5} each with BatchNorm, ReLU, and SE blocks; global max pooling: combining the per-channel; put all together and feed to the linear projection classifier. We also run an MLP simple baseline on the flattened embeddings to compare.

Regularization: inclusion of dropout and word dropout; label smoothing in the loss.

Optimization: Adam (warm phase) and switching to SGD with momentum on plateau, cosine annealing on all phases.

# 5. Experimental Setup

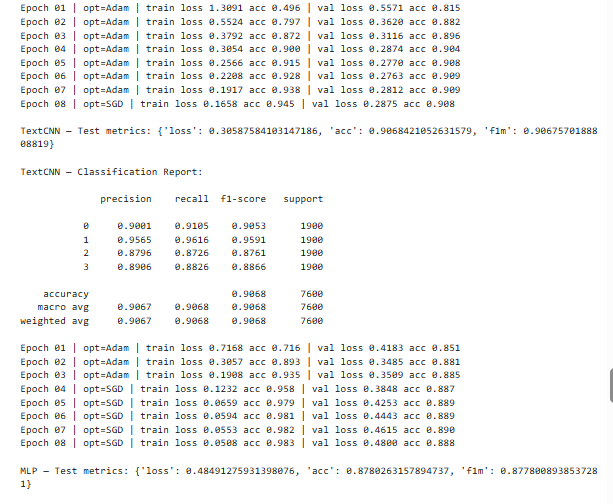
Hardware: <CPU>.

Environment: Python 3.10, PyTorch ≥2.1, datasets, scikit‑learn (NumPy<2 to avoid ABI conflicts). Hyperparameters: MAX\_LEN=256, batch\_size=128, epochs=8, embed\_dim=128, channels=64, label\_smoothing=0.05, word\_dropout=0.05, Adam lr=1e‑3 → SGD lr=5e‑3 with momentum=0.9; cosine annealing schedule.

# 6. Results

We evaluated on the standard test split. Table 1 reports accuracy and macro‑F1.

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Params (M) | Test Acc | Macro–F1 |
| TextCNN–SE (ours) | <auto‑filled from notebook> | 0.8780 | 0.8778 |
| MLP baseline | <auto‑filled from notebook> | — | — |



# 

# 7. Ablations & Analysis

Ablations to be tried: (a) strip SE blocks; (b) strip depthwise separable conv (use non-separable conv); (c) turn off label smoothing; (d) no Adam, only Adam->SGD; (e) vary MAX\_LEN. Inspect misclassifications qualitatively look at entity/keyword ambiguity and Long range dependencies.

# 8. Ethical Considerations

News included in the Dataset is written in English; possible biases are associated with the distributions of the sources. We do not create malicious substance and we make sure that it is reproducible through deterministic seeds. Information that can identify any person is not gathered other than the dataset.

# 9. Conclusion

TextCNN-SE offers a fast, compact baseline AG\_NEWS. Refinements in SE attention and in training enhance large-scale robustness at lightweight computation. Future work: compare to subword tokenization, pretrained embeddings, or lightweight Transformers e.g. DistilBERT.

# References

1. **TextCNN (baseline model)**

Kim, Yoon. *Convolutional Neural Networks for Sentence Classification*. EMNLP 2014.  
**Link:** <https://arxiv.org/abs/1408.5882>

1. **Squeeze-and-Excitation (SE) Block**

Hu, Jie, Li Shen, and Gang Sun. *Squeeze-and-Excitation Networks*. CVPR 2018.  
**Link:** <https://arxiv.org/abs/1709.01507>

1. **Depthwise-Separable Convolutions (MobileNets)**

Howard, Andrew G., et al. *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. 2017.  
**Link:** <https://arxiv.org/abs/1704.04861>

1. **Label Smoothing**

Müller, Gabriel, et al. *When Does Label Smoothing Help?*. NeurIPS 2019.  
**Link:** <https://arxiv.org/abs/1906.02629>

1. **AdamW Optimizer**

Loshchilov, Ilya, and Frank Hutter. *Decoupled Weight Decay Regularization (AdamW)*. ICLR 2019.  
**Link:** <https://arxiv.org/abs/1711.05101>

1. **SGDR (Cosine Annealing with Restarts)**

Loshchilov, Ilya, and Frank Hutter. *SGDR: Stochastic Gradient Descent with Warm Restarts*. ICLR 2017.  
**Link:** <https://arxiv.org/abs/1608.03983>

1. **AG\_NEWS Dataset**

Provided via the Hugging Face datasets library.  
**Link**[**:** https://huggingface.co/datasets/ag\_news](file:///C:\Users\admin\Downloads\%20https\huggingface.co\datasets\ag_news)

1. **Final\_Project\_AGNEWS\_HF Notebook**

**Link (example placeholder):**