KvochurHegel at NakbaArchiveClassifier Shared Task: Nakba Image Classification via ConvNeXt-V2 and Label Smoothing

Minh-Hoang Le
University of Information Technology, Ho Chi Minh City, Vietnam
Vietnam National University, Ho Chi Minh City, Vietnam
24520542@gm.uit.edu.vn

Abstract

This paper presents the KvochurHegel team's submission to the Nakba Image Classification shared task at the Nakba-NLP 2026 Workshop. The task requires the binary classification of social media images into `destruction` and `not_destruction` categories. Given a limited and imbalanced training set of 1,400 images, we utilized a ConvNeXt-V2 Nano backbone combined with extensive data augmentation and label smoothing, prioritizing standard regularization over task-specific architectural modifications. For inference, we applied a 6-view Test-Time Augmentation (TTA) strategy using a hard-voting mechanism. The baseline system achieved a Macro F1-score of 0.8593 and an Accuracy of 0.8706 on the official private test set, ranking 6th out of 16 participating teams.

Keywords: Image Classification, Destruction Detection, Social Media Imagery, Conflict Imagery, ConvNeXt-V2

1. Introduction

The Nakba Image Classification shared task (Abrahams et al., 2026), part of the Nakba-NLP 2026 Workshop, focuses on categorizing social media imagery to improve the accessibility of archived records from the Tech for Palestine incubator project. The dataset comprises 2,001 images sourced from Instagram, published by Palestinian content creators and journalists in Gaza between October 7, 2023, and December 15, 2025. The objective is to develop a binary image classifier capable of distinguishing images that show destroyed or damaged infrastructure (`destruction`) from those that do not (`not_destruction`).

Our team, KvochurHegel, prioritized a standard, regularized Convolutional Neural Network (CNN) over task-specific architectural modifications, given the dataset's class imbalance (64.7% / 35.3%) and the visual noise common in social media imagery. We submitted a ConvNeXt-V2 Nano baseline trained with extensive data augmentation and evaluated using a multi-view Test-Time Augmentation (TTA) strategy. The system ranked 6th out of 16 participating teams, achieving a Macro F1-score of 0.8593 on the official private test set.

2. Experimental Setup

To support reproducibility, all training and inference experiments described in Section 3 were conducted in the following software environment:

- Python 3.10.0
- PyTorch 2.9.1 (compiled with CUDA 12.6)
- MMEngine 0.10.7
- MMCV 2.2.0
- MMPretrain 1.2.0

3. Methodology

The Nakba Image Classification task provides a dataset of 2,001 images, partitioned into 1,400 for training, 199 for validation, and 402 for testing. The training set consists of 906 `not_destruction` and 494 `destruction` images (64.7% / 35.3%). We use the validation set to select the model checkpoint that achieves the highest Macro F1-score, which is the primary evaluation metric for the shared task. All experiments are conducted on a single NVIDIA RTX 4060 GPU with 8GB of VRAM.

3.1. Model Architecture

We use the ConvNeXt-V2 Nano backbone, selected for offering a competitive accuracy-to-parameter ratio, as demonstrated in (Woo et al., 2023). We initialize the backbone with weights pre-trained on ImageNet-21K and fine-tuned on ImageNet-1K at 384×384 resolution. The classification head is a linear layer mapping the 640-dimensional global average pooled features to the two target classes.

3.2. Training and Regularization

The model is trained for 50 epochs with a batch size of 16 using the AdamW optimizer (lr $= 2 \times 10^{-4}$, weight decay$= 0.3$) (Loshchilov and Hutter, 2019). We adopt this configuration from the MMPreTrain baseline, which was used during the backbone's original ImageNet-1K fine-tuning phase. The learning rate follows a 5-epoch linear warmup and a cosine annealing schedule decaying to 1×10^{-6}.

As a regularizer for the small training set, we apply a Label Smoothing (Müller et al., 2019) coefficient of 0.2, determined on the validation set. Training augmentations include *RandomResized-Crop*, horizontal flips, and the `timm_increasing`

Proceedings of the 2nd International Workshop on Nakba Narratives as Language Resources (Nakba-NLP 2026) @ LREC 2026, pages 118–120
11 May 2026.

RandAugment policy (Cubuk et al., 2020). We also apply *RandomErasing* ($p = 0.25$) (Zhong et al., 2020) to discourage the model from over-fitting to local patches. We maintain an Exponential Moving Average (EMA) of the weights with a smoothing factor of 0.9995 (implemented as 0.0005 momentum in MMPreTrain (Contributors, 2023)) to obtain a weight-averaged checkpoint less sensitive to late-training oscillations. Given the constrained one-week development timeline, no explicit class-weighting or oversampling was implemented to address the 64.7% / 35.3% dataset imbalance.

3.3. Inference Strategy

We use a 6-view Test-Time Augmentation (TTA) strategy for the final evaluation. Test images are resized to 384 pixels on the short edge. We then extract three 384×384 spatial crops (left/top, center, and right/bottom) sampled along the longer dimension, along with their horizontally mirrored counterparts.

Final labels are assigned through a hard-voting mechanism. For each of the 6 views, a vote is cast for `destruction` if the softmax probability for that class exceeds 0.5. The majority class is selected as the final prediction. In the case of a tie (3 votes each), the prediction from the original center-cropped view serves as the tie-breaker.

4. Results

Our team, KvochurHegel, submitted our ConvNeXt-V2 baseline to the Nakba Image Classification shared task. The system was evaluated on the private test set of 402 images.

The system achieved a final Macro F1-score of 0.8593 and an Accuracy of 0.8706. Table 1 reports all evaluation metrics returned by the task organizers. The parity between Macro Recall (0.8616) and Precision (0.8572) suggests that the model did not exhibit a systematic bias toward the majority class, despite the lack of explicit class-weighting during training. Additionally, the identical values for Recall, Specificity, and Balanced Accuracy (0.8616) indicate that the model achieved equal per-class recalls across both the minority and majority classes.

Our submission ranked 6th out of 16 participating teams. The winning system achieved a Macro F1-score of 0.899. The margin of 0.0397 F1 to the first-place entry suggests that a standard, regularized CNN pipeline performed competitively on this task.

Metric	**Score**
Macro F1-score	0.8593
Accuracy	0.8706
Balanced Accuracy	0.8616
Precision	0.8572
Recall	0.8616
Specificity	0.8616

Table 1: Official private test set evaluation metrics for the KvochurHegel team submission.

5. Conclusion

We described a ConvNeXt-V2 Nano baseline system developed for the Nakba Image Classification shared task. To address the constraints of a small 1,400-image training set and a 64.7% / 35.3% class imbalance, the approach relied on adopting the MMPreTrain default hyperparameters for ConvNeXt-V2, extensive data augmentation, label smoothing, and a 6-view Test-Time Augmentation (TTA) strategy. These design choices prioritized standard regularization over task-specific architectural modifications.

The system achieved a Macro F1-score of 0.8593, ranking 6th out of 16 teams. This suggests that a standard, regularized CNN pipeline performs competitively on this task, finishing within 0.04 F1 of the winning system. Due to a constrained one-week development timeline, explicit class-aware training mechanisms were omitted. We leave the exploration of oversampling and class-weighted loss functions on this dataset to future work.

6. Bibliographical References

Alexei Abrahams, Shadi Abudalfa, Mustafa Jarrar, and George Mikros. 2026. The nakbaarchive-classifier shared task on nakba image classification. In *Proceedings of the 2nd International Workshop on Nakba Narratives as Language Resources (Nakba-NLP 2026), co-located with the Language Resources and Evaluation Conference (LREC 2026)*, Palma, Mallorca, Spain.

MMPreTrain Contributors. 2023. Openmmlab's pre-training toolbox and benchmark. `https://github.com/open-mmlab/mmpretrain`.

Ekin D. Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V. Le. 2020. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. In *International Conference on Learning Representations*.

Rafael Müller, Simon Kornblith, and Geoffrey E Hinton. 2019. When does label smoothing help?

In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Sanghyun Woo, Shoubhik Debnath, Ronghang Hu, Xinlei Chen, Zhuang Liu, In So Kweon, and Saining Xie. 2023. Convnext v2: Co-designing and scaling convnets with masked autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16133–16142.

Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. 2020. Random erasing data augmentation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(07):13001–13008.