# Supplementary Material for Exemplar-Free Incremental Deepfake Detection

Paper #1078

**Table 1**: Data split and collected years for the datasets in the D-IDD and T-IDD protocols.

| Protocol | Dataset | Year | Train | Val | Test |
|---|---|---|---|---|---|
| D-IDD | FF++ | 2019 | 180k | 35k | 35k |
| | DFD | 2019 | - | - | 170K |
| | Celeb-DF | 2020 | 301k | 25k | 26K |
| | DFDC-P | 2020 | 188k | 23k | 39K |
| | DFFD | 2020 | 65k | 8k | 17K |
| | FFIW | 2021 | 788k | 25k | 172K |
| | OpenForensics | 2021 | 105K | 15k | 30K |
| | ForgeryNet | 2021 | 460K | 87k | 121K |
| | Kodf | 2021 | - | - | 75K |
| | ForgeryNIR | 2022 | 45K | 5k | 10K |
| T-IDD | DeepFakes | 2018 | 73k | 15k | 23K |
| | FS-GAN | 2019 | 67k | 11k | 14K |
| | SC-FEGAN | 2019 | 64K | 13k | 19K |
| | DF-StarGAN | 2019 | 58K | 8k | 10K |
| | StyleGAN2 | 2020 | 70k | 16k | 15K |
| | BlendFace | 2020 | 62k | 10k | 14K |
| | MaskGAN | 2020 | 66k | 14k | 26K |
| | FaceShifter | 2020 | - | - | 37K |
| | StarGAN2 | 2020 | - | - | 29k |

## 1 Appendix

In the appendix, we introduce the detailed inference pipeline of the proposed method, more experimental details and results.

### *1.1 Algorithm Details.*

**Inference pipeline.** Figure 1 illustrates the inference pipeline of the proposed method. For the inference phase, we suggest the following steps: 1) feeding the given test image into the $F_e$ to obtain the image feature $f$, 2) calculating similarity to search for the nearest domain center for the given test image, 3) Image features $f$ are input to $F_s$, and intermediate features are blended with prompts from adapters associated with the nearest domain center, 4) Feed the output from $F_s$ to the classifier related to the nearest domain center to obtain the final prediction.

### *1.2 Experimental Details*

**Datasets.** In the task of Exemplar-Free Incremental Deepfake Detection (EF-IDD), we build the D-IDD and T-IDD protocols utilizing diverse deepfake dataset: FF++ , Celeb-DF [6], DFDC-P [2],DFFD [1], FFIW [9],OpenForensics [5], ForgeryNIR [8], ForgeryNet [4]. The data split and collected years of these datasets are presented in Table 1. Note that we build T-IDD using ten subsets from ForgeryNet. Both D-IDD and T-IDD simulate possible practical scenarios of the Deepfake detection problem. The D-IDD protocol contains eight sessions of altogether 4.77 million samples, and the T-IDD protocol contains eight sessions of 0.81 million samples.

**Evaluation metrics.** Let $S_{i,t}$ be the evaluation score, $e.g.$, classification accuracy on the $i$-th task after training on the $t$-th task. After the model finishes training on the $t$-th task, we compute the Average Accuracy (AA) and Average Forgetting (AF) as follows:

$$AA = \frac{1}{t}\sum_{i=1}^{t} S_{i,t}, \tag{1}$$

$$AF = \frac{1}{t-1}\sum_{i=1}^{t-1} {}_{j\in\{1,\cdots,t-1\}} \left(S_{i,j} - S_{i,t}\right). \tag{2}$$

Note that Average Accuracy is the overall evaluation metric for continual learning, which includes two aspects: greater learning capacity and less catastrophic forgetting, while Average Forgetting only serves as a measure of catastrophic forgetting.

**Complexity Analysis.** We analyze floating-point operations (FLOPs) to compare the complexity of different EF-DIL methods. For fair comparison, we use ViT-B/16 [3] as the backbone network. As shown in Figure 2, FT, EWC, SI and LwF represent normal ViT computations. Our method only supplements an additional adapter (containing three lightweight convolutional layers) with only a slight increase in computational effort. Based on ViT, existing prompt-based query key mechanisms or clustering strategies are designed to automatically select relevant hints for each instance respectively. These strategies require the instance to be fed into the network twice and involve additional computational overhead. Specifically, the prompts interacts with image tokens through a multi-head attention mechanism in the forward process to acquire domain-specific knowledge. However, the computational complexity of its self-attention is quadratic to the length of the input sequence [7]. Therefore, increasing the number of hints results in additional computational overhead. In contrast, our approach can perform computations in a structured and sparse manner, achieving the best balance of performance and computational efficiency.

## References


[1] H. Dang, F. Liu, J. Stehouwer, X. Liu, and A. K. Jain. On the detection of digital face manipulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5781–5790, 2020.

[2] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and C. C. Ferrer. The deepfake detection challenge (dfdc) preview dataset. *ArXiv*, 2019.

[3] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021.

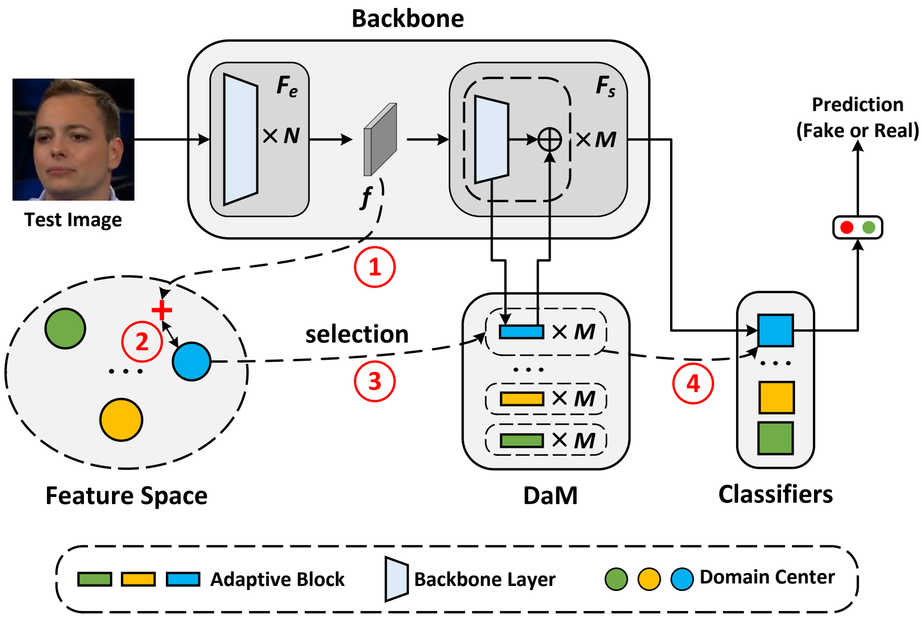


**Figure 1**: Illustration of the inference pipeline of the proposed method. The displayed indexes correspond to the following four inference steps respectively: 1) obtaining the feature of a given test image, 2) searching for the nearest domain center obtained by performing $softmax$ operation on similarity between the feature and domain centers 3) feeding the feature $f$ into $F_s$ where intermediate features are multi-stage blended with prompts from the adapters associated nearest domain center, 4) obtain the final prediction by the classifier associated with the nearest domain center.

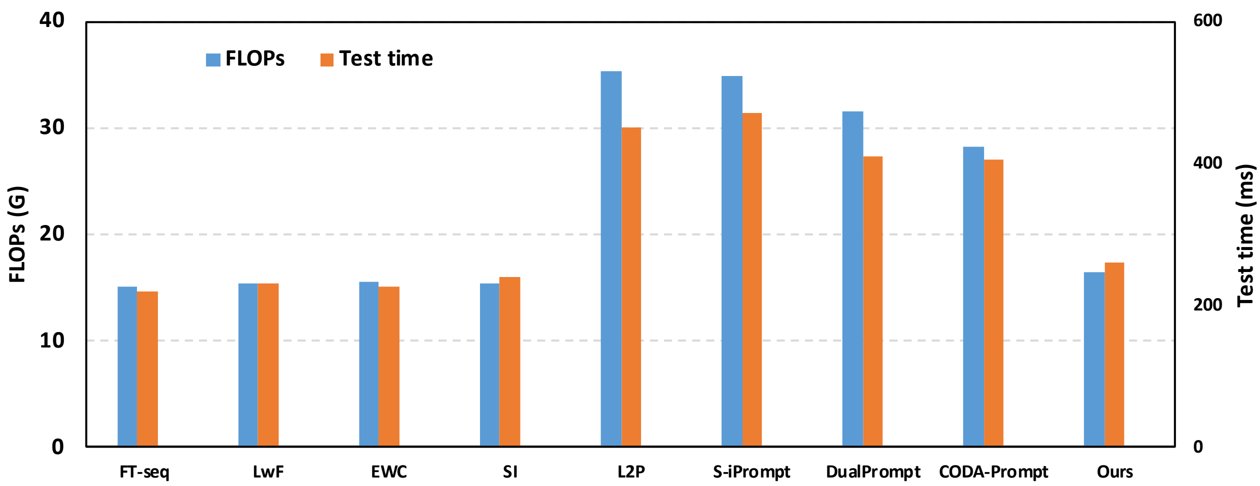


**Figure 2**: Computational cost of different EF-DIL methods.

**Table 2**: Performance of the proposed EF-IDD framework under D-IDD and T-IDD protocols. AA and AF are calculated for each session.

| Data-incremental Deepfake detection (D-IDD) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| sessions | FF++ | Celeb-DF | DFDC-P | DFFD | FFIW | OpenForensics | ForgeryNIR | ForgeryNet | AA (↑) | AF (↓) |
| FF++ | 95.84 | - | - | - | - | - | - | - | 95.84 | - |
| Celeb-DF | 94.27 | 75.51 | - | - | - | - | - | - | 84.89 | 1.57 |
| DFDC-P | 92.79 | 74.26 | 76.08 | - | - | - | - | - | 81.04 | 2.14 |
| DFFD | 91.04 | 73.84 | 73.22 | 78.35 | - | - | - | - | 79.10 | 3.10 |
| FFIW | 89.82 | 72.95 | 71.69 | 74.27 | 77.64 | - | - | - | 77.27 | 4.26 |
| OpenForensics | 87.68 | 72.45 | 70.65 | 73.41 | 73.55 | 80.34 | - | - | 76.48 | 5.12 |
| ForgeryNIR | 86.61 | 72.26 | 70.43 | 72.05 | 71.62 | 73.28 | 77.32 | - | 74.71 | 6.33 |
| ForgeryNet | 86.54 | 72.18 | 70.11 | 71.25 | 69.29 | 70.13 | 68.57 | 69.35 | 71.59 | 7.12 |
| Type-incremental Deepfake detection (T-IDD) | | | | | | | | | | |
| sessions | DeepFakes | StyleGAN2 | FS-GAN | BlendFace | MaskGAN | SC-FEGAN | DF-StarGAN | DiscoFaceGAN | AA (↑) | AF (↓) |
| DeepFakes | 80.84 | - | - | - | - | - | - | - | 80.84 | - |
| StyleGAN2 | 78.59 | 76.44 | - | - | - | - | - | - | 77.51 | 2.25 |
| FS-GAN | 77.24 | 73.16 | 77.28 | - | - | - | - | - | 75.89 | 3.46 |
| BlendFace | 76.83 | 72.13 | 73.24 | 77.59 | - | - | - | - | 74.94 | 4.09 |
| MaskGAN | 75.37 | 71.43 | 72.62 | 72.25 | 76.32 | - | - | - | 73.61 | 5.14 |
| SC-FEGAN | 73.76 | 71.03 | 70.96 | 71.36 | 70.27 | 76.52 | - | - | 72.32 | 6.21 |
| DF-StarGAN | 72.58 | 70.53 | 70.65 | 70.81 | 70.12 | 69.29 | 79.51 | - | 71.64 | 7.02 |
| DiscoFaceGAN | 71.64 | 69.15 | 69.32 | 70.25 | 69.06 | 67.83 | 67.22 | 72.46 | 69.83 | 8.54 |


[4] Y. He, B. Gan, S. Chen, Y. Zhou, G. Yin, L. Song, L. Sheng, J. Shao, and Z. Liu. Forgerynet: A versatile benchmark for comprehensive forgery analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4360–4369, 2021.

[5] T.-N. Le, H. H. Nguyen, J. Yamagishi, and I. Echizen. Openforensics: Large-scale challenging dataset for multi-face forgery detection and segmentation in-the-wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10117–10127, 2021.

[6] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu. Celeb-df: A large-scale challenging dataset for deepfake forensics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3207–3216, 2020.

[7] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10012–10022, 2021.

[8] Y. Wang, C. Peng, D. Liu, N. Wang, and X. Gao. Forgerynir: deep face forgery and detection in near-infrared scenario. *IEEE Transactions on Information Forensics and Security*, 17:500–515, 2022.

[9] T. Zhou, W. Wang, Z. Liang, and J. Shen. Face forensics in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5778–5788, 2021.