Graph-Attention Network with Adversarial Domain Alignment for Robust Cross-Domain Facial Expression Recognition

Razieh Ghaedi* ROZAGHAEDI90@GMAIL.COM
Manchester Metropolitan University

AmirReza BabaAhmadi AMIRREZABABAAHMADI@GMAIL.COM
University of Tehran

Reyer Zwiggelaar RRZ@ABER.AC.UK
Aberystwyth University

Xinqi Fan X.FAN@MMU.AC.UK

Nashid Alam N.ALAM@MMU.AC.UK
Manchester Metropolitan University

Editors: Hung-yi Lee and Tongliang Liu

Abstract

Cross-domain facial expression recognition (CD-FER) remains difficult due to severe domain shift between training and deployment data. We propose Graph-Attention Network with Adversarial Domain Alignment (GAT-ADA), a hybrid framework that couples a ResNet-50 as backbone with a batch-level Graph Attention Network (GAT) to model inter-sample relations under shift. Each mini-batch is cast as a sparse ring graph so that attention aggregates cross-sample cues that are informative for adaptation. To align distributions, GAT-ADA combines adversarial learning via a Gradient Reversal Layer (GRL) with statistical alignment using CORAL and MMD. GAT-ADA is evaluated under a standard unsupervised domain adaptation protocol: training on one labeled source (RAF-DB) and adapting to multiple unlabeled targets (CK+, JAFFE, SFEW 2.0, FER2013, and ExpW). GAT-ADA attains 74.39% mean cross-domain accuracy. On RAF-DB→FER2013, it reaches 98.0% accuracy, corresponding to a ≈36-point improvement over the best baseline we re-implemented with the same backbone and preprocessing.

Keywords: Cross-Domain Learning, Unsupervised Domain Adaptation, Graph Attention Networks, Gradient Reversal.

1. Introduction

Facial Expression Recognition (FER) is focused on automating human emotion detection (Guo and Zhang, 2019). FER enhances human-computer interaction (Jaimes and Sebe, 2007), healthcare monitoring (Bisogni et al., 2022) and affective computing (Poria et al., 2017). While recent deep learning advances have achieved remarkable performance in controlled laboratory settings, real-world deployment remains severely constrained by domain shift, the fundamental problem where models trained on one dataset fail catastrophically when applied to data from different demographic populations, imaging conditions, or cultural contexts (Wang et al., 2024). This challenge is particularly acute in FER, where subtle variations in lighting, camera quality, ethnic representation, and cultural expression

© 2025 R. Ghaedi, A. BabaAhmadi, R. Zwiggelaar, X. Fan & N. Alam.

norms can render even state-of-the-art models ineffective (Gopalan et al., 2014). To address these limitations, Cross-Domain Facial Expression Recognition (CD-FER) is developed as a strategy to enhance cross-domain adaptation capability across diverse datasets (Li and Deng, 2022b). However, current CD-FER methods face two critical limitations. First, conventional CNN pipelines overlook relational structure that is useful for adaptation, particularly inter-sample relations across heterogeneous domains (Kopalidis et al., 2024). Second, existing domain adaptation techniques rely primarily on marginal distribution alignment, overlooking fine-grained statistical discrepancies between source and target domains. These limitations result in typical cross-domain accuracies below 60%, despite achieving over 90% within training domains. Graph-based learning offers a complementary perspective by modeling relationships with attention on graphs. In FER, graphs are typically built within a single face (e.g., landmarks/regions), with models like the Graph Attention Network (GAT) (Veličković et al., 2018) over landmark graphs or related relation-aware designs (Prados-Torreblanca et al., 2022; Ma and Ma, 2023). However, such approaches only look inside one image and do not capture connections between different samples, which can be very useful for domain adaptation (Prados-Torreblanca et al., 2022; Ma and Ma, 2023).

Our proposed Graph-Attention Network with Adversarial Domain Alignment (GAT-ADA) addresses this gap. GAT-ADA couples a ResNet-50 backbone for robust intra-image feature extraction with a batch-level GAT that models inter-sample relationships under domain shift. Concretely, each mini-batch is cast as a sparse graph (ring connectivity), and the GAT attends across samples to emphasize cross-sample similarities that are informative for adaptation. To align source and target distributions, we combine adversarial learning via a Gradient Reversal Layer (GRL) (Ganin et al., 2016) with statistical alignment using CORAL (Sun and Saenko, 2016) and MMD (Gretton et al., 2012). GAT-ADA achieves strong CD-FER performance through three key contributions:

1. Hybrid Graph-Based Feature Modeling: A hybrid approach integrating GATs with ResNet-50 is proposed to dynamically model inter-sample relationships (by treating each image as a graph node) while preserving deep hierarchical feature extraction.

2. Multi-Component Domain Adaptation: Domain adaptation is improved by incorporating CORAL and MMD losses alongside a GRL for adversarial adaptation. This facilitates more effective feature alignment across different datasets.

3. Comprehensive Benchmark Evaluation: The proposed model is systematically evaluated across six benchmark datasets (ExpW, RAF-DB, CK+, FER2013, JAFFE, and SFEW2.0) to assess its effectiveness in CD-FER.

2. Image Databases

A robust evaluation of CD-FER models requires datasets that reflect the complexity and diversity of facial expressions in varied contexts. In this study, we use unsupervised domain adaptation (UDA) with one labeled source and multiple unlabeled targets. We train on labeled RAF-DB (Li et al., 2017b) as the source; unlabeled CK+ (Lucey et al., 2010), JAFFE (Lyons et al., 1998), SFEW 2.0 (Dhall et al., 2015), FER2013 (Goodfellow et al., 2013), and ExpW (Zhang et al., 2017) as targets. All datasets are mapped to the common

7 class set (Anger, Disgust, Fear, Happiness, Sadness, Surprise, Neutral). These datasets span both in-the-wild and lab-controlled environments and differ in terms of image distribution factors such as lighting conditions, ethnic and cultural representation, image resolution and quality, subject age, gender balance, and expression authenticity. This diversity establishes a challenging and representative test-bed for CD-FER, enabling us to assess the model's ability to adapt to varied and unseen conditions, which is an essential criterion for robust real-world deployment. Figure 1 shows examples from the six datasets.

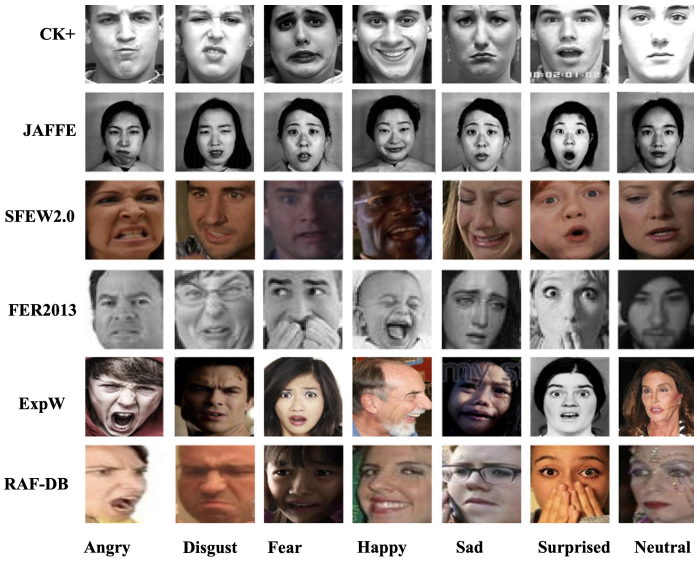

Figure 1: Facial expression examples.

2.1. Data Preprocessing

All experiments employ a standardized data pipeline to reduce domain discrepancies. We first resize every face image to a fixed 100×100 resolution to ensure consistent input dimensions. Then we apply data augmentation; specifically random horizontal flips (50% probability) and slight color jitter (brightness ±0.1) to mimic variations in pose and lighting conditions. Finally, the augmented images are converted to tensors and normalized (each channel mean=0.5, std=0.5) for numerical stability.

3. Methodology

3.1. Model Architecture

GAT-ADA is a hybrid pipeline with three main components: (i) a ResNet-50 backbone that produces compact 512-D embeddings, (ii) a relational module implemented as a batch-level Graph Attention Network (GAT) operating on batch-level image embeddings to model inter-sample dependencies, and (iii) a multi-component domain adaptation head that combines a task classifier with adversarial alignment (GRL) and statistical alignment (CORAL, MMD). An overview is shown in Fig. 2.

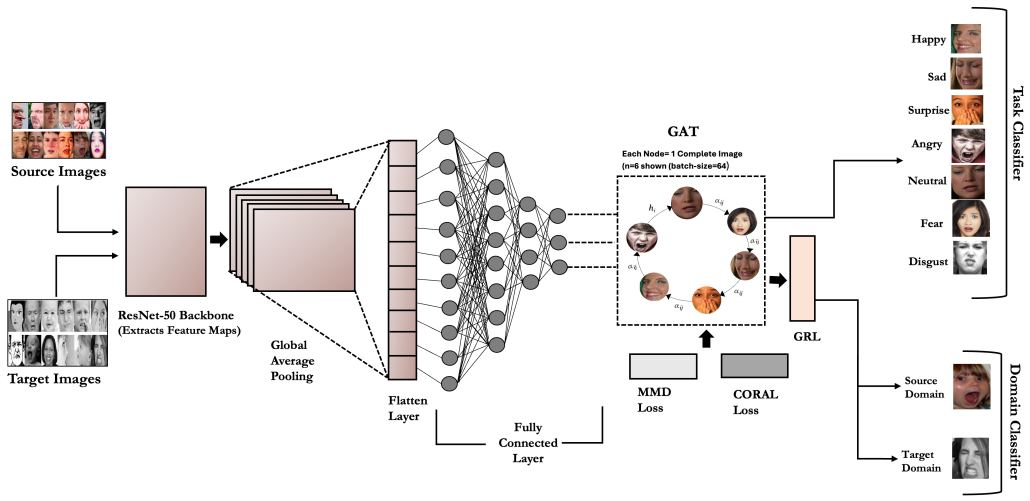

Figure 2: Pipeline of the proposed GAT-ADA framework.

3.2. Feature Extraction via ResNet-50

To adapt ResNet-50 for CD-FER, we remove the final classification layer and extract deep feature maps. Each input image $x^{(i)}$ is encoded as

$$F_{\text{initial}}^{(i)} = \text{ResNet50}\left(x^{(i)}\right) \in \mathbb{R}^{2048\times h\times w}, \tag{1}$$

followed by Global Average Pooling (GAP) to reduce spatial dimensions:

$$F_{\text{pooled}}^{(i)} = \text{GAP}\left(F_{\text{initial}}^{(i)}\right) \in \mathbb{R}^{2048}. \tag{2}$$

A subsequent linear projection with nonlinearity yields a 512-D embedding:

$$h_i \;=\; \text{ReLU}\big(W_{\text{proj}}\, F_{\text{pooled}}^{(i)} + b_{\text{proj}}\big) \in \mathbb{R}^{512}, \tag{3}$$

where $W_{\text{proj}} \in \mathbb{R}^{512\times 2048}$ and $b_{\text{proj}} \in \mathbb{R}^{512}$ are trainable. These embeddings $\{h_i\}$ retain localized emotional cues while providing compact inputs for subsequent relational modeling and domain alignment.

3.3. Graph-Based Relational Modeling with GAT

Unlike conventional GAT approaches that treat facial regions or landmarks as nodes, GAT-ADA constructs graphs at the batch level, with each image represented as a node. This design targets inter-sample relationships across domains, crucial for domain adaptation but not the primary focus of intra-image landmark graphs or feature self-attention, which operate within a single face or within a single embedding, respectively. In our framework, the ResNet-50 backbone supplies strong intra-image representations, while the GAT learns a relational structure between samples, weighting embeddings by cross-sample similarity or dissimilarity.

Connectivity choice: We adopt a ring topology to enforce structured information flow with $\mathcal{O}(n)$ edges (versus $\mathcal{O}(n^2)$ for fully connected graphs), providing a favorable accuracy–stability–efficiency trade-off. Mini-batches are randomly shuffled each epoch, so ring neighborhoods change over iterations, acting as stochastic neighbor sampling across the dataset. In contrast, landmark graphs (intra-image) and feature self-attention (within-embedding) operate within a single sample and thus do not explicitly capture inter-sample relationships, the focus of our batch-level GAT.

Single-image inference: During training and at inference time when mini-batches contain multiple images, we use a ring graph. When only a single test image is available, the ring graph cannot be constructed. We therefore use a prototype memory bank: representative prototypes are stored during training, and a query q forms a star-topology graph by connecting to its top-k prototypes using cosine similarity:

$$\mathrm{sim}(q,p) = \frac{q^\top p}{\|q\|\|p\|}, \qquad \mathcal{N}_k(q) = \text{top-}k_{\,p\in\mathcal{P}}\,\mathrm{sim}(q,p), \qquad E_q = \{(q,p)\,|\,p \in \mathcal{N}_k(q)\}. \tag{4}$$

GAT attention and aggregation are then applied over $\mathcal{N}_k(q)$, preserving relational benefits in single-image settings.

Formal definition. For a mini-batch of size n, nodes are samples and edges follow a circular pattern:

$$E = \{(i,j) \mid i \in [0, n-1],\ j = (i+1) \bmod n\} \cup \{(j,i) \mid (i,j) \in E\}, \tag{5}$$

with adjacency $A \in \mathbb{R}^{n\times n}$ given by

$$A_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Let $\mathcal{N}_i = \{\, j \mid A_{ij} = 1 \,\}$ denote the neighbor set of node i. Let $h_i \in \mathbb{R}^{512}$ be the projected feature of sample i defined in Eq. (3). Multi-head attention with K heads computes:

$$\alpha_{ij}^k = \frac{\exp(\sigma(a_k^\top [W_k h_i \,||\, W_k h_j]))}{\sum_{l\in\mathcal{N}_i} \exp(\sigma(a_k^\top [W_k h_i \,||\, W_k h_l]))}, \tag{7}$$

where $W_k \in \mathbb{R}^{512\times 512}$ is the weight matrix for head k, $a_k \in \mathbb{R}^{1024}$ is the attention parameter vector, $||$ denotes concatenation, σ is LeakyReLU with negative slope 0.2, and $k \in \{1,\ldots,K\}$ indexes the K attention heads. Node updates are:

$$h_i' = \sigma\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j\in\mathcal{N}_i} \alpha_{ij}^k\, W_k h_j\right), \tag{8}$$

which enhances cross-domain alignment by updating each sample's embedding with an attention-weighted combination of its neighbors.

3.4. Multi-Component Domain Alignment

To bridge distributional gaps between the source and target domains, GAT-ADA incorporates a hybrid domain adaptation framework that combines adversarial alignment with statistical distribution matching. GAT-ADA follows an UDA paradigm, where labeled source domain data (RAF-DB) and unlabeled target domain data are both available during training. This approach ensures that the learned feature representations are both emotion-discriminative and domain-invariant, enabling robust adaptation across datasets with different acquisition conditions, demographics, and lighting variations. At the core of the adversarial component is a domain classifier that receives the 512-dimensional encoded features through a Gradient Reversal Layer (GRL). The GRL acts as an identity function during forward propagation but multiplies gradients by $-\lambda_{\text{GRL}}$ during backpropagation:

$$\text{GRL}(x) = \begin{cases} x & \text{(forward pass)} \\ -\lambda_{\text{GRL}} \frac{\partial L}{\partial x} & \text{(backward pass)} \end{cases} \tag{9}$$

This forces the encoder to learn domain-invariant features by confusing the domain classifier. The domain classifier is trained with binary cross-entropy loss separately for source and target batches. The task classification loss for emotion recognition on the source domain is defined as:

$$L_{\text{task}} = -\frac{1}{n_s} \sum_{i=1}^{n_s} \sum_{c=1}^{7} y_i^c \log \hat{y}_i^c, \tag{10}$$

where n_s is the number of source samples, y_i^c is the ground-truth label (one-hot encoded for 7 emotion classes), and $\hat{y}_i^c$ is the predicted probability for class c. To complement adversarial alignment, we introduce two statistical alignment losses:

1. **CORAL loss** aligns second-order statistics by minimizing the Frobenius norm between the covariance matrices of source and target domains:

$$L_{\text{CORAL}} = \frac{1}{4d^2} \left\| C_s - C_t \right\|_F^2, \tag{11}$$

   where C_s and C_t are the covariance matrices of source and target features, and d is the feature dimension.

2. **MMD loss** aligns first-order statistics by minimizing the mean shift:

$$L_{\text{MMD}} = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(f_i^s) - \frac{1}{n_t} \sum_{j=1}^{n_t} \phi(f_j^t) \right\|^2, \tag{12}$$

   where $\phi(\cdot)$ denotes the kernel mapping to a Reproducing Kernel Hilbert Space (RKHS) using a Gaussian RBF kernel.

The total objective balances classification and alignment losses:

$$L_{\text{total}} = L_{\text{task}} + L_{\text{domain}_s} + L_{\text{domain}_t} + \lambda_{\text{align}} \big(L_{\text{CORAL}} + L_{\text{MMD}} \big), \tag{13}$$

where L_{domain_s} and L_{domain_t} are the binary cross-entropy losses of the domain classifier with GRL for source and target batches, respectively. We set $\lambda_{\text{align}} = 1.0$ and $\lambda_{\text{GRL}} = 1.0$ in all experiments, which yielded stable training without oscillations. This three-pronged domain adaptation strategy captures both marginal and statistical distribution discrepancies. CORAL emphasizes covariance structure alignment, while MMD addresses global mean matching. Together with the adversarial signal, this synergy enables GAT-ADA to achieve robust CD-FER performance.

3.5. Training Configuration

We train the model for 32 epochs using the AdamW optimizer to improve adaptation. The initial learning rate is set to 0.0001. To stabilize training, we used a two-phase schedule: a linear warm-up for 5 epochs followed by cosine annealing decay over the remaining 27 epochs. This approach prevents unstable updates and ensures fine-tuned convergence. We evaluated performance using accuracy, precision, recall, F1 score, and AUC. The specific parameter values used in our experiments for training the model are summarized in Table 1.

Table 1: Meta-parameter information for the network in Figure 2

Parameter	**Value**
Epochs	32
Optimizer	AdamW
Initial learning rate	10^{-4}
Learning rate schedule	5-epoch warm-up; cosine annealing for 27 epochs
Classification loss	Cross-entropy
Adversarial loss	Binary cross-entropy (with logits) with GRL
Statistical alignment	CORAL + MMD ($\lambda_{\text{align}} = 1.0$)
GRL gradient scaling	$\lambda_{\text{GRL}} = 1.0$
GAT attention heads	4
GAT LeakyReLU negative slope	0.2
Batch size	64

4. Results

Domain adaptation performance is highly sensitive to backbone architecture and dataset composition Chen et al. (2022). To ensure fair comparison, we adopt the AGRA benchmarking protocol, which standardizes backbone architectures and dataset splits across all competing methods. This addresses the inconsistency in previous evaluations where different studies used varying experimental setups. As shown in Table 2, GAT-ADA achieves strong cross-domain performance. With a ResNet-50 backbone, the method attains a mean accuracy of 74.39%, outperforming the previous best FER-DAS by 4.79 percentage points under our re-implementation. Notably, on RAF-DB→FER2013, we observe up to 98.04% accuracy under our protocol, a notable improvement over the strongest baseline with the same backbone and preprocessing. With the lighter ResNet-18, GAT-ADA still achieves 70.28% mean accuracy, demonstrating scalability and efficiency.

Table 2: Comparison of accuracies of state-of-the-art methods in CD-FER using ResNet-50 and ResNet-18 backbones. RAF-DB is used as the source domain; CK+, JAFFE, SFEW2.0, FER2013, and ExpW are target domains.

Method	**Backbone**	**CK+**	**JAFFE**	**SFEW2.0**	**FER2013**	**ExpW**	**Mean**
DFA Zhu et al. (2016)	ResNet-50	64.26	44.44	43.07	45.79	56.86	50.88
LPL Li et al. (2017a)	ResNet-50	74.42	53.05	48.85	55.89	66.90	59.82
FTDNN Zavarez et al. (2017)	ResNet-50	79.07	52.11	47.48	55.98	67.72	60.47
DETN Li and Deng (2018)	ResNet-50	78.22	55.89	49.40	52.29	47.58	56.68
CADA Long et al. (2018)	ResNet-50	72.09	52.11	53.44	57.61	63.15	59.68
ICID Ji et al. (2019)	ResNet-50	74.42	50.70	48.85	53.70	69.54	59.44
SAFN Xu et al. (2019)	ResNet-50	75.97	61.03	52.98	55.64	64.91	62.11
SWD Xu et al. (2019)	ResNet-50	75.19	54.93	52.06	55.84	68.35	61.27
ETD Li et al. (2020)	ResNet-50	75.16	51.19	52.77	50.41	67.82	59.47
ECAN Li and Deng (2022a)	ResNet-50	79.77	57.28	52.29	56.46	47.37	58.63
JUMBOT Fatras et al. (2021)	ResNet-50	79.46	54.13	51.97	53.56	63.69	60.56
AGRA Chen et al. (2022)	ResNet-50	85.27	61.50	56.43	58.95	68.50	66.13
AGLRLS Gao et al. (2024)	ResNet-50	87.60	61.97	58.28	60.68	73.0	68.30
DCD-DAN Alzahrani et al. (2025)	ResNet-50	-	62.68	64.76	-	78.37	68.60
USTST Guo et al. (2024)	ResNet-50	84.83	51.49	53.44	-	55.98	60.44
FER-DAS Zhu et al. (2025)	ResNet-50	89.25	69.93	57.41	61.8	-	69.60
GAT-ADA (2025)	**ResNet-50**	**88.04**	**54.06**	**58.14**	**98.04**	**73.65**	**74.39**
DFA Zhu et al. (2016)	ResNet-18	54.26	42.25	38.30	47.88	47.42	46.02
LPL Li et al. (2017a)	ResNet-18	72.87	53.99	49.31	53.61	68.35	59.63
FTDNN Zavarez et al. (2017)	ResNet-18	76.74	50.23	49.54	53.28	68.08	59.57
CADA Long et al. (2018)	ResNet-18	73.64	55.40	52.29	54.71	63.74	59.96
DETN Li and Deng (2018)	ResNet-18	64.19	52.11	42.25	42.01	43.92	48.90
ICID Ji et al. (2019)	ResNet-18	67.44	48.83	47.02	53.00	68.52	56.96
SAFN Xu et al. (2019)	ResNet-18	68.99	49.30	50.46	53.31	68.32	58.08
SWD Xu et al. (2019)	ResNet-18	72.09	53.52	49.31	50.58	61.45	58.00
ECAN Li and Deng (2022a)	ResNet-18	66.51	52.11	48.21	50.76	48.73	53.26
ETD Li et al. (2020)	ResNet-18	72.34	49.44	49.67	47.66	64.62	56.75
AGRA Chen et al. (2022)	ResNet-18	79.84	61.03	51.15	51.95	65.03	61.80
JUMBOT Fatras et al. (2021)	ResNet-18	76.67	52.10	49.19	50.58	61.45	58.00
CSRL Li et al. (2022)	ResNet18	88.37	66.67	-	55.53	-	-
GAT-ADA (2025)	**ResNet-18**	**81.06**	**48.06**	**53.48**	**98.68**	**70.11**	**70.28**

Table 3: Comprehensive performance metrics for GAT-ADA on each target dataset

Dataset	Backbone	Prec.	Recall	F1	AUC
CK+	ResNet-50	87.09	87.42	87.22	0.9993
JAFFE	ResNet-50	54.16	54.18	53.71	0.9992
SFEW 2.0	ResNet-50	58.26	58.44	57.92	0.9994
FER2013	ResNet-50	98.06	98.04	98.04	0.9804
ExpW	ResNet-50	73.58	73.63	73.56	0.9995
Mean	**ResNet-50**	**74.22**	**74.34**	**74.09**	**0.9993**
CK+	ResNet-18	80.46	80.72	80.52	0.9989
JAFFE	ResNet-18	48.22	48.46	47.61	0.9993
SFEW 2.0	ResNet-18	53.52	53.66	53.04	0.9998
FER2013	ResNet-18	98.68	98.68	98.68	0.9919
ExpW	ResNet-18	70.09	70.12	70.05	0.9993
Mean	**ResNet-18**	**70.19**	**70.32**	**69.98**	**0.9993**

4.1. Computational Efficiency Analysis

We evaluated GAT-ADA's efficiency against recent state-of-the-art CD-FER methods. Table 4 reports FLOPs, latency, memory, and training time, measured on an NVIDIA Tesla T4 (16 GB) with PyTorch in evaluation mode.

Table 4: Computational efficiency comparison of GAT-ADA with recent state-of-the-art methods in CD-FER.

Method	Params (M)	FLOPs (G)	Inference (ms)	Memory (MB)	Training (min/epoch)
GAT-ADA	**14.11**	**1.77**	**2.80**	**180**	**0.82**
AGLRLS	-	-	780.00	-	-
AGRA	25.30	6.77	20.84	200	11.88
DCD-DAN	25.60	6.77	15.20	195	8.45
USTST	-	-	13.56	-	-

GAT-ADA delivers the strongest accuracy–cost balance: it achieves lower FLOPs, memory, and latency than recent baselines while maintaining state–of–the–art accuracy. It processes an image in 2.8 ms on a Tesla T4 ($\approx$ 357 FPS), supporting real–time throughput on our setup and showing lower latency than AGRA, DCD–NAN, USTST, and AGLRLS in Table 4. Training is also faster per epoch and requires fewer resources. These results demonstrate that our hybrid architecture achieves strong performance without sacrificing computational practicality, making GAT-ADA suitable for resource-constrained environments and real-time applications. As a visual complement to Table 4, Figure 3 summarizes the accuracy–efficiency trade-offs of GAT-ADA versus recent CD-FER methods.

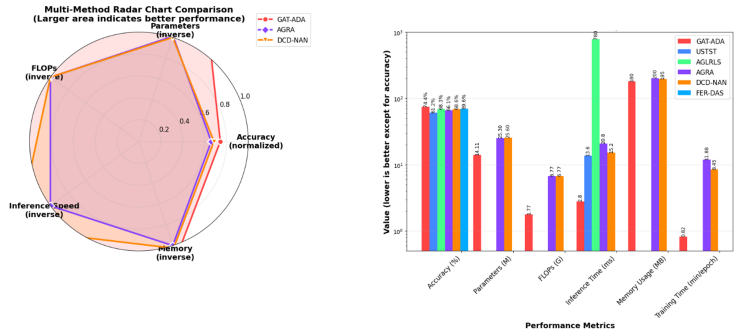

Figure 3: Performance-efficiency summary across methods. Left: radar chart with cost axes inverted (larger area indicates a better overall profile). Right: bar chart of individual metrics. GAT-ADA shows the strongest balance, high mean accuracy with markedly lower FLOPs, latency, and memory supporting real-time.

4.2. Confusion Matrix Analysis

We analyzed the classification performance of our proposed framework by examining the confusion matrices for both ResNet-18 and ResNet-50 backbones. These results highlight key insights into classification strengths, misclassification trends, and model robustness in handling domain shifts across different datasets. Figure 4 and Figure 5 illustrate the confusion matrices for ResNet-50 and ResNet-18, showing classification performance when transferring knowledge from RAF-DB to FER2013 and CK+ datasets. To ensure reproducibility, we fixed random seeds during training for each dataset and model configuration. As a result, the misclassified cases in the confusion matrices remain consistent across runs, unless major changes are made to the architecture or hyperparameters. Additional confusion matrices for other dataset transfers, including RAF-DB to ExpW, JAFFE, and SFEW 2.0 with ResNet-18 and ResNet-50, are provided in Supplementary Material.

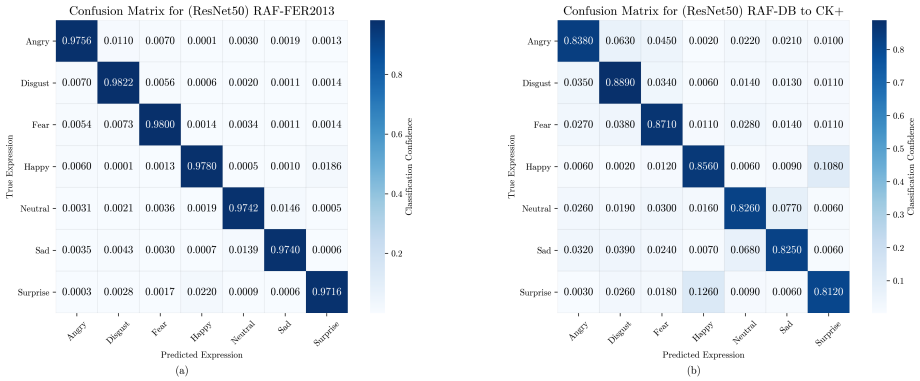

Figure 4: Confusion matrices for ResNet-50: (a) RAF-DB to FER2013, (b) RAF-DB to CK+

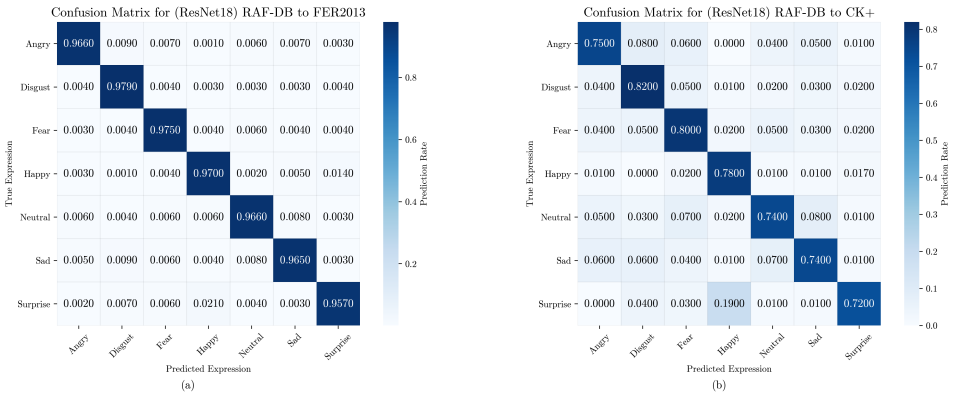

Figure 5: Confusion matrices for ResNet-18: (a) RAF-DB to FER2013, (b) RAF-DB to CK+

5. Discussion

Our empirical results show notable improvements over existing approaches (Table 2),with GAT-ADA achieving 98.04% accuracy on the challenging RAF-DB to FER2013 transfer task, 36 percentage point improvement over existing methods. This achieves notable improvement over existing approaches stems from three key innovations: First, our novel batch-level graph construction treats each image as a node, enabling dynamic modeling of inter-sample relationships that captures expression pattern similarities across domains. Second, our multi-component domain adaptation strategy (combining MMD, CORAL, and GRL) provides comprehensive alignment of both marginal and statistical distributions. Third, the synergistic integration of these components creates a robust framework that handles extreme domain shifts between color and grayscale images with varying resolutions and contexts. While GAT-ADA achieves notable improvement over existing approaches on large-scale transfers, certain datasets present unique challenges. JAFFE poses exceptional difficulties with only 213 images from 10 subjects, creating extreme domain shifts and minimal sample diversity. In such settings, the limited variability between samples reduces the relational information available to the graph attention mechanism, which can constrain performance. We have employed a prototype memory bank strategy for single-image inference to preserve relational information in the absence of batch-level context. As a potential future extension, we plan to adapt this mechanism for training on extremely small datasets by enabling cross-batch prototype accumulation and progressive refinement of class representations. This would allow the model to benefit from aggregated feature diversity over multiple iterations, reducing the sensitivity of GAT-ADA to small sample variability and enhancing robustness under severe data scarcity. Similarly, SFEW 2.0's low resolution and noisy video-derived labels constrain domain alignment effectiveness. While GAT-ADA remains competitive and outperforms recent methods on SFEW 2.0, these results highlight that certain datasets may require tailored adaptations beyond our current framework. Our method demonstrates optimal effectiveness in large-scale cross-domain settings with sufficient diversity, as evidenced by exceptional FER2013 results and strong CK+ and ExpW performance.

5.1. Ablation Studies

To evaluate the effectiveness of each architectural component, we conducted comprehensive ablation studies. Table 5 presents performance metrics across different model configurations, revealing several critical insights.

Impact of Graph Attention Mechanisms: The comparison between GAT and Graph Convolutional Network (GCN) Kipf and Welling (2016) implementations reveals the crucial role of attention mechanisms in CD-FER. As shown in Table 5, Replacing GAT with GCN causes a substantial drop (e.g., 98.0% → 61.0%), indicating that attention over inter-sample neighbors is crucial for CD-FER.

Domain Alignment: The impact of GRL on model performance is particularly noteworthy. As evidenced in Table 5, removing the GRL significantly degrades performance from 98.04% to 64.51% accuracy in the ResNet-50 + GAT configuration. This substantial performance gap (33.53 percentage points) validates our hypothesis that robust domain alignment is crucial for cross-domain adaptation. The GRL's effectiveness stems from its ability to enforce domain-invariant feature learning while preserving emotion-discriminative information.

Behavioral Analysis of ResNet-50 as a Backbone Architecture: An unexpected yet significant finding emerged in our backbone architecture comparison. Despite ConvNeXt being more modern, its coupling with batch-level GAT under our protocol trails the ResNet-50 backbone, suggesting that compatibility with inter-sample relational modeling and alignment is more critical than raw feature capacity.

While ResNet-18 exhibited stronger performance in a specific transfer scenario (RAF-DB → FER2013), our broader evaluation across all datasets (Table 2) confirms ResNet-50's superior mean accuracy. Based on this comprehensive assessment, we selected the hybrid configuration (ResNet-50 + GAT + GRL) as our optimal setup, striking a balance between feature extraction depth, domain adaptation robustness, and classification accuracy. The chosen model achieves outstanding results across all metrics, 98.04% accuracy, 98.06% precision, 98.04% recall, and 0.988 AUC, demonstrating the synergistic benefits of combining attention mechanisms with domain adaptation strategies.

Table 5: Comparison of performance metrics across different model configurations.

Model Configuration	**Accuracy**	**Precision**	**Recall**	**F1-score**	**AUC**
ResNet-50 + GAT + GRL (✓)	0.9804	0.9806	0.9804	0.9804	0.9804
ResNet-50 + GAT + GRL (✗)	0.6451	0.6451	0.6451	0.6164	0.7564
ResNet-50 + GCN + GRL (✗)	0.4690	0.5000	0.4690	0.3751	0.5855
ResNet-50 + GCN + GRL (✓)	0.6104	0.6055	0.6104	0.5642	0.7141
ConvNext + GAT + GRL (✗)	0.6353	0.6691	0.6353	0.5892	0.7368
ConvNext + GAT + GRL (✓)	0.9674	0.9681	0.9674	0.9672	0.9799
ResNet-18 + GAT + GRL (✗)	0.6551	0.6551	0.6551	0.6264	0.7654
ResNet-18 + GAT + GRL (✓)	0.9868	0.9868	0.9868	0.9868	0.9919

6. Conclusion

In this work, we addressed the limitations of current CD-FER methods by proposing GAT-ADA, a framework that integrates GAT with ResNet-50 and a multi-component domain adaptation strategy combining MMD, CORAL, and GRL. Evaluated under a unified benchmark with consistent source/target datasets and feature extractors, GAT-ADA demonstrated superior effectiveness in handling real-world variability in FER. Our batch-level graph construction, which treats each image as a node to model inter-sample relationships, proved particularly effective for cross-domain emotion recognition, while the multi-component alignment strategy successfully captured both marginal and statistical distribution discrepancies between source and target domains. While GAT-ADA was specifically designed and validated for FER, the core architectural principles of relational modeling through graph attention and hybrid domain adaptation represent general design patterns that could potentially be adapted to other cross-domain learning problems. However, such extensions would require domain-specific graph construction strategies, appropriate connectivity patterns, and thorough empirical validation in each target domain. Future work will systematically explore this transferability by conducting preliminary experiments in medical imaging and action recognition, improving computational efficiency through graph sparsification, enhancing robustness on small or noisy datasets, and extending to multimodal emotion recognition. These investigations will help determine whether GAT-ADA's design principles can serve as a general paradigm for domain-adaptive learning across diverse applications, or whether the observed success is primarily specific to the structured nature of facial expression data.

References

Ahmed Omar Alzahrani, Ahmed Mohammed Alghamdi, M. Usman Ashraf, Iqra Ilyas, Nadeem Sarwar, Abdulrahman Alzahrani, and Alaa Abdul Salam Alarood. A novel facial expression recognition framework using deep learning based dynamic cross-domain dual attention network. *PeerJ Comput. Sci.*, 11:e2866, 2025. doi: 10.7717/peerj-cs.2866.

Carmen Bisogni, Aniello Castiglione, Sanoar Hossain, Fabio Narducci, and Saiyed Umer. Impact of deep learning approaches on facial expression recognition in healthcare industries. *IEEE Trans. Ind. Informatics*, 18(8):5619–5627, 2022. doi: 10.1109/TII.2022.3141400.

Tianshui Chen, Tao Pu, Hefeng Wu, Yuan Xie, Lingbo Liu, and Liang Lin. Cross-domain facial expression recognition: A unified evaluation benchmark and adversarial graph learning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(12):9887–9903, 2022. doi: 10.1109/TPAMI.2021.3131222.

Abhinav Dhall, Roland Goecke, Simon Lucey, and Tom Gedeon. Static facial expressions in the wild 2.0 (sfew2.0) dataset. Selected static frames from the AFEW video dataset as part of the EmotiW2015 challenge, 2015.

Kilian Fatras, Thibault Séjourné, Rémi Flamary, and Nicolas Courty. Unbalanced minibatch optimal transport: Applications to domain adaptation. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, pages 3186–3197, 2021.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor S. Lempitsky. Domain-adversarial training of neural networks. *J. Mach. Learn. Res.*, 17(59):1–35, 2016.

Yuefang Gao, Yuhao Xie, Zeke Zexi Hu, Tianshui Chen, and Liang Lin. Adaptive global-local representation learning and selection for cross-domain facial expression recognition. *IEEE Trans. Multim.*, 26:6676–6688, 2024. doi: 10.1109/TMM.2024.3355637.

Ian J. Goodfellow, David Erhan, Pierre-Luc Carrier, Aaron Courville, and Yoshua Bengio. The facial expression recognition 2013 (fer2013) dataset. Introduced as part of "Challenges in Representation Learning: A report on three machine learning contests" at ICML2013, 2013.

Raghuraman Gopalan, Ruonan Li, and Rama Chellappa. Unsupervised adaptation across domain shifts by generating intermediate data representations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 36(11):2288–2302, 2014. doi: 10.1109/TPAMI.2013.249.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander J. Smola. A kernel two-sample test. *J. Mach. Learn. Res.*, 13:723–773, 2012.

Guodong Guo and Na Zhang. A survey on deep learning based face recognition. *Comput. Vis. Image Underst.*, 189, 2019. doi: 10.1016/J.CVIU.2019.102805.

Zhe Guo, Bingxin Wei, Jiayi Liu, Xuewen Liu, Zhibo Zhang, and Yi Wang. USTST: unsupervised self-training similarity transfer for cross-domain facial expression recognition. *Multim. Tools Appl.*, 83(14):41703–41723, 2024. doi: 10.1007/s11042-023-17362-9.

Alejandro Jaimes and Nicu Sebe. Multimodal human-computer interaction: A survey. *Comput. Vis. Image Underst.*, 108(1-2):116–134, 2007. doi: 10.1016/J.CVIU.2006.10.019.

Yanli Ji, Yuhan Hu, Yang Yang, Fumin Shen, and Heng Tao Shen. Cross-domain facial expression recognition via an intra-category common feature and inter-category distinction feature fusion network. *Neurocomputing*, 333:231–239, 2019. doi: 10.1016/j.neucom.2018.12.037.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *CoRR*, abs/1609.02907, 2016.

Thomas Kopalidis, Vassilios Solachidis, Nicholas Vretos, and Petros Daras. Advances in facial expression recognition: A survey of methods, benchmarks, models, and datasets. *Inf.*, 15(3):135, 2024. doi: 10.3390/INFO15030135.

Mengxue Li, Yi-Ming Zhai, You-Wei Luo, Peng-Fei Ge, and Chuan-Xian Ren. Enhanced transport distance for unsupervised domain adaptation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13936–13944, 2020. doi: 10.1109/CVPR42600.2020.01395.

Shan Li and Weihong Deng. Deep emotion transfer network for cross-database facial expression recognition. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 3092–3099, 2018. doi: 10.1109/ICPR.2018.8545284.

Shan Li and Weihong Deng. A deeper look at facial expression dataset bias. *IEEE Transactions on Affective Computing*, 13(2):881–893, 2022a. doi: 10.1109/TAFFC.2020.2973158.

Shan Li and Weihong Deng. A deeper look at facial expression dataset bias. *IEEE Trans. Affect. Comput.*, 13(2):881–893, 2022b. doi: 10.1109/TAFFC.2020.2973158.

Shan Li, Weihong Deng, and JunPing Du. Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2852–2861, July 2017a.

Shan Li, Weihong Deng, and JunPing Du. The real-world affective faces database (raf-db). Crowdsourced dataset from "Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild," presented at CVPR2017, 2017b.

Yingjian Li, Zheng Zhang, Bingzhi Chen, Guangming Lu, and David Zhang. Deep margin-sensitive representation learning for cross-domain facial expression recognition. *IEEE Transactions on Multimedia*, 25:1359–1373, 2022. doi: 10.1109/TMM.2022.3141604.

Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I. Jordan. Conditional adversarial domain adaptation. In *Advances in Neural Information Processing Systems*, volume 31, pages 1647–1657, 2018. doi: 10.5555/3326943.3327094.

Patrick Lucey, Jeffrey F. Cohn, Takeo Kanade, Jason M. Saragih, Zara Ambadar, and Iain A. Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 94–101, 2010. doi: 10.1109/CVPRW.2010.5543262.

Michael J. Lyons, Shigeru Akamatsu, Miyuki G. Kamachi, and Jiro Gyoba. The japanese female facial expression (jaffe) database. Website, originally presented alongside the paper "Coding Facial Expressions with Gabor Wavelets" at the Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, April 14–16,1998, 1998.

Guanming Ma and Huazhu Ma. Relation-aware network for facial expression recognition. In *17th IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, 2023.

Soujanya Poria, Erik Cambria, Rajiv Bajpai, and Amir Hussain. A review of affective computing: From unimodal analysis to multimodal fusion. *Inf. Fusion*, 37:98–125, 2017. doi: 10.1016/J.INFFUS.2017.02.003.

Andrés Prados-Torreblanca, José M. Buenaposada, and Luis Baumela. Shape preserving facial landmarks with graph attention networks. In *British Machine Vision Conference (BMVC)*, 2022.

Baochen Sun and Kate Saenko. Deep CORAL: correlation alignment for deep domain adaptation. In Gang Hua and Hervé Jégou, editors, *Computer Vision - ECCV 2016 Workshops - Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III*, volume 9915 of *Lecture Notes in Computer Science*, pages 443–450, 2016. doi: 10.1007/978-3-319-49409-8_35.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 – May 3, 2018, Conference Track Proceedings*, 2018.

Yan Wang, Shaoqi Yan, Yang Liu, Wei Song, Jing Liu, Yang Chang, Xinji Mai, Xiping Hu, Wenqiang Zhang, and Zhongxue Gan. A survey on facial expression recognition of static and dynamic emotions. *CoRR*, abs/2408.15777, 2024. doi: 10.48550/ARXIV.2408.15777.

Ruijia Xu, Guanbin Li, Jihan Yang, and Liang Lin. Larger norm more transferable: An adaptive feature norm approach for unsupervised domain adaptation. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1426–1435, 2019. doi: 10.1109/ICCV.2019.00151.

Marcus Vinicius Zavarez, Rodrigo Ferreira Berriel, and Thiago Oliveira-Santos. Cross-database facial expression recognition based on fine-tuned deep convolutional network. In *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 405–412. IEEE, 2017.

Zhanpeng Zhang, Ping Luo, Chen Change Loy, and Xiaoou Tang. Expression in-the-wild (expw) dataset. Introduced in "From Facial Expression Recognition to Interpersonal Relation Prediction," Int. J. Comput. Vis. (2018), 2017.

Ronghang Zhu, Gaoli Sang, and Qijun Zhao. Discriminative feature adaptation for cross-domain facial expression recognition. In *2016 International Conference on Biometrics (ICB)*, pages 1–7, 2016.

Yanan Zhu, Jiaqiu Ai, Weibao Xue, Mingyang Wu, Sen Yang, Wei Jia, and Min Hu. Cross-domain facial expression recognition: Bi-Directional fusion of active and stable information. *Eng. Appl. Artif. Intell.*, 149:110357, 2025. doi: 10.1016/j.engappai.2024.110357.