HalfMix Augmentation and Regularized Dual-Path Learning for Cross-Domain Gaze Estimation

Jiuk Hong and Heechul Jung
Department of Artificial Intelligence, Kyungpook National University
{hong4497, heechul}@knu.ac.kr

KYUNGPOOK NATIONAL UNIVERSITY
BMVC 2025

Github

INTRODUCTION & MOTIVATION

Introduction

Cross-domain gaze estimation addresses the challenge of training robust gaze estimation models that can generalize effectively across different datasets without requiring access to target domain data during training.

Key Challenges

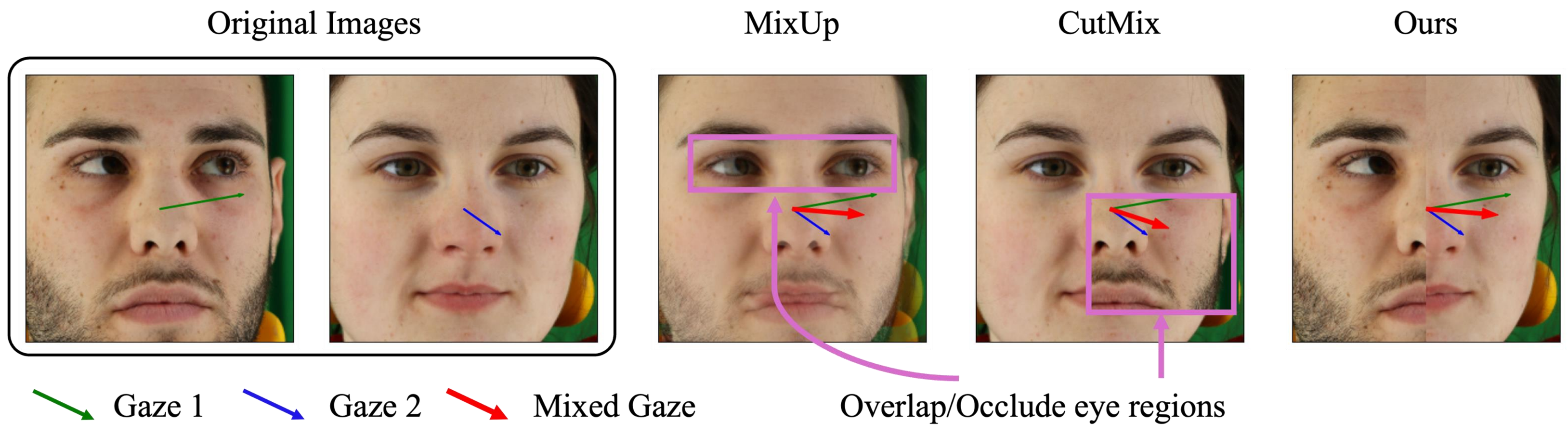
- **Limited generalization:** Current models tend to overfit to source domain-specific characteristics such as subject identity and imaging conditions, resulting in poor transferability to unseen target domains
- **Feature entanglement:** Existing approaches struggle to disentangle gaze-relevant features from domain-specific confounding factors like subject appearance and environmental conditions

Motivation

Existing augmentation methods damage eye region integrity:

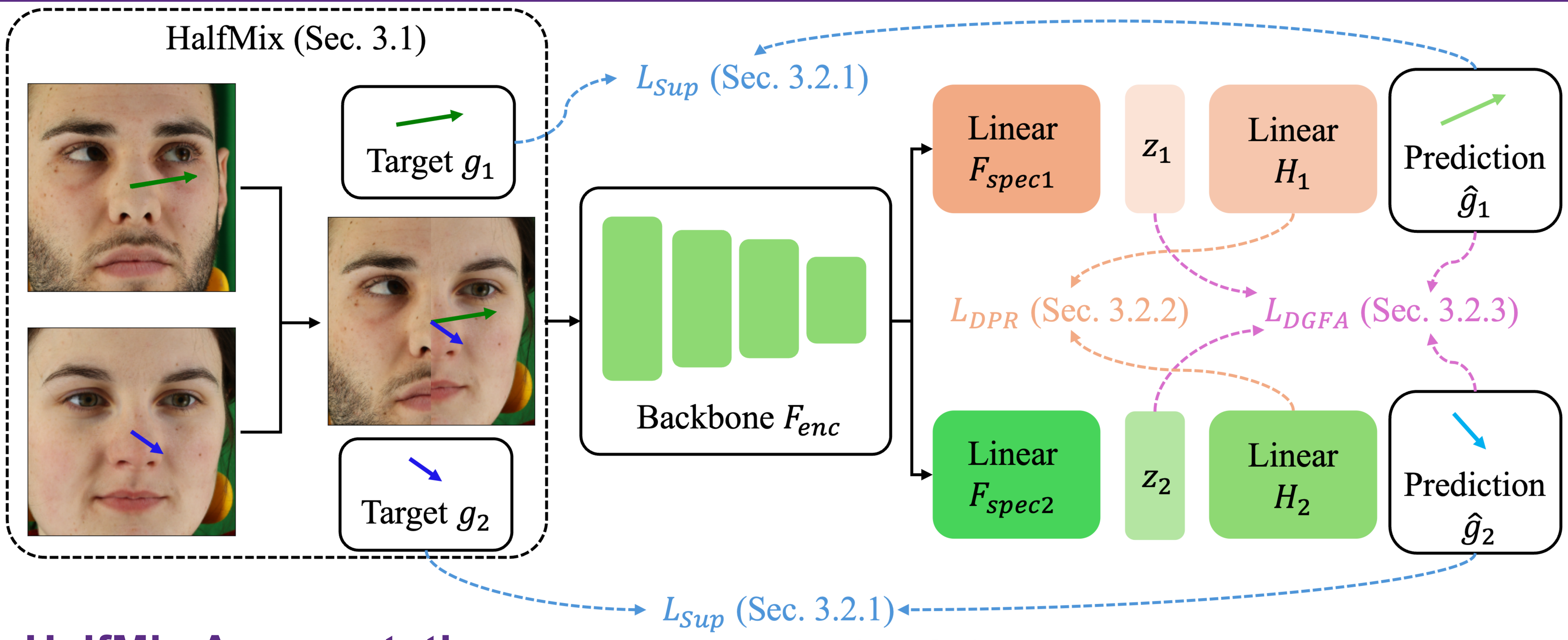
Problems

- **MixUp:** Overlaps eye regions, loses clarity
- **CutMix:** Occludes critical eye features
- **Neither preserves bilateral eye structure**



Our Solution: HalfMix preserves both eye regions while enabling effective augmentation.

METHOD



HalfMix Augmentation

$$I_{mix}(x, y) = \begin{cases} I_a(x, y), & if\ y < W/2 \\ I_b(x, y), & if\ y \geq W/2 \end{cases}$$

- Splits two images vertically at center
- Combines left half of image A with right half of image B
- Both eye regions remain intact and unoccluded

Dual-Path Learning

- Mixed gaze labels from augmented images are ambiguous for a single pathway, so dual pathways separately predict each source's gaze direction
- Ensemble predictions at inference improve robustness

Loss Details

L_{Sup}: Supervised Loss

Standard gaze direction prediction loss for both pathways

L_{DPR}: Diversity-Promoting

Forces pathways to focus on different information by diversifying their learned parameters

$$L_{DPR} = CosSim(W_{H_1}, W_{H_2}) + KL(W_{H_1}, W_{H_2})$$

L_{DGFA}: Dual-Gaze Alignment

Aligns features from both halves for gaze consistency

$$L_{DGFA} = MSE(Sim_{feat}(z_1, z_2), Sim_{gaze}(g_1, g_2))$$

$$L_{total} = L_{Sup} + \alpha \cdot L_{DPR} + \beta \cdot L_{DGFA}$$

EXPERIMENTAL RESULTS

Baseline Comparison

Method	D_E	$D_E \to D_M$	$D_E \to D_D$	D_G	$D_G \to D_M$	$D_G \to D_D$
ResNet18	4.98	7.34	8.13	14.55	12.81	10.46
Proposed	5.60 (-12.5%)	6.69 (8.9%)	7.97 (2.0%)	13.36 (8.2%)	8.74 (31.8%)	8.43 (19.4%)
ResNet50	4.61	7.62	8.02	14.69	14.06	12.41
Proposed	5.52 (-19.7%)	6.82 (10.5%)	6.99 (12.8%)	16.11 (-9.7%)	9.06 (35.6%)	9.14 (26.3%)

- Major generalization to unseen domains, though in-domain accuracy slightly decreases as a trade-off for robustness

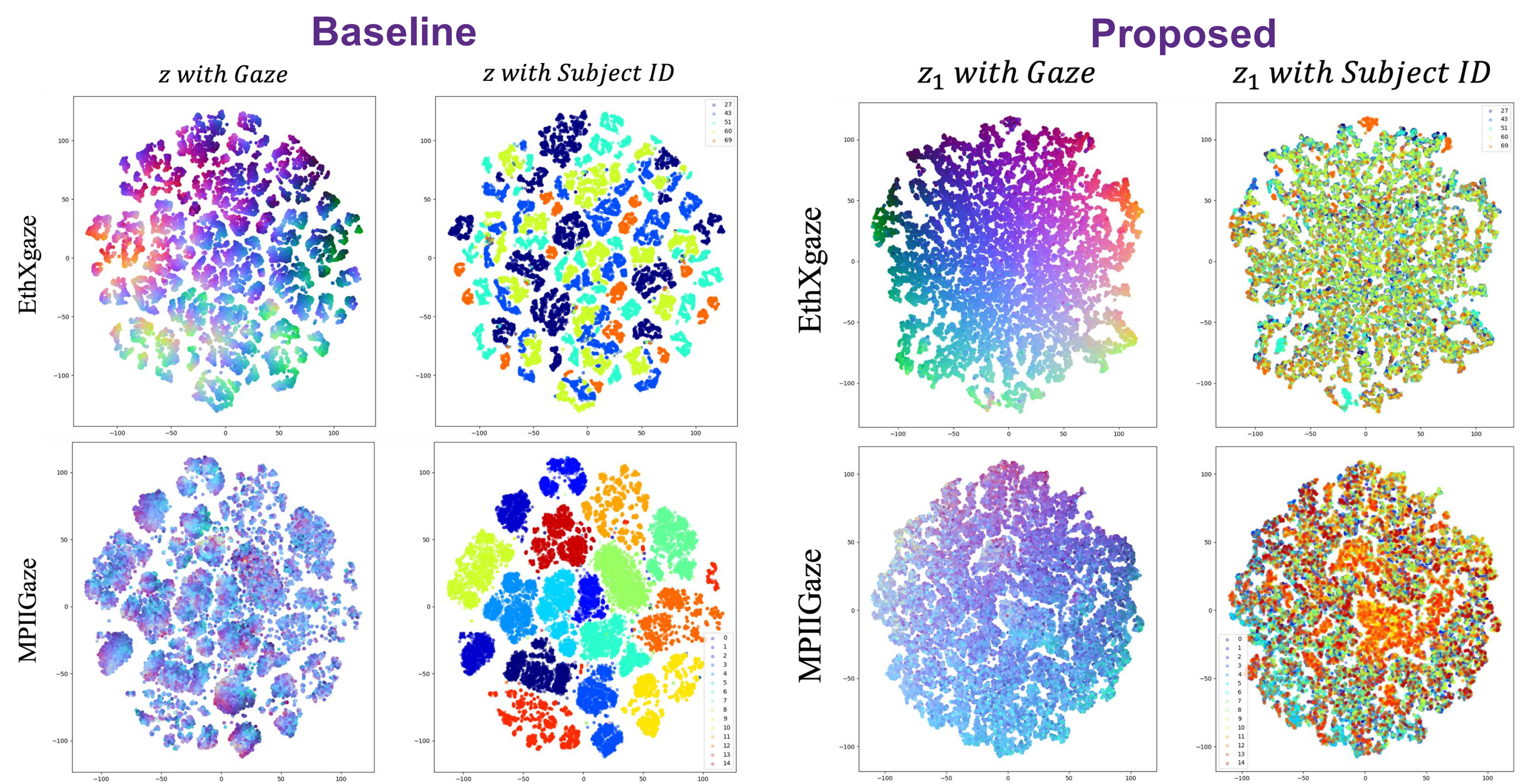
SOTA Comparison

Method	TD	$D_E \to D_M$	$D_E \to D_D$	$D_G \to D_M$	$D_G \to D_D$
GazeAdv [13]	✓	-	-	8.19	12.27
PnP-GA [11]	✓	6.00	6.17	5.74	7.04
CRGA [14]	✓	5.48	5.66	5.89	6.49
LatentGaze [10]		7.98	9.81	-	-
PureGaze [4]		7.08	7.48	9.28	9.32
AGG [1]		7.10	7.07	7.87	7.93
GazeConsistent [15]		**6.50**	7.44	7.55	**7.03**
CDG [14]		6.73	7.95	**7.03**	7.27
Ours (ResNet18)		6.69	8.31	8.74	8.43
Ours (ResNet50)		6.82	**6.99**	9.06	9.14

- We achieve 6.99° on EyeDiap, establishing new state-of-the-art performance among target-data-free methods that require no adaptation data.

Feature Analysis

t-SNE shows clusters by **subject ID**, ours by **gaze direction**.



- Unlike baseline features that memorize subject appearance, our method extracts generalizable gaze cues independent of identity.

Conclusion

We propose HalfMix augmentation and dual-path learning with DPR and DGFA regularization for robust cross-domain gaze estimation. Our method achieves state-of-the-art target-data-free performance by learning domain-invariant gaze features that cluster by direction rather than identity, though with a minor trade-off in source domain accuracy.