FaceGCD: Generalized Face Discovery via Dynamic Prefix Generation

Yunseok Oh[1, 2], Dong-Wan Choi[1]
Department of Computer Science and Engineering, Inha University, South Korea[1]
AutoLabs, Inc., South Korea[2]

Source Code

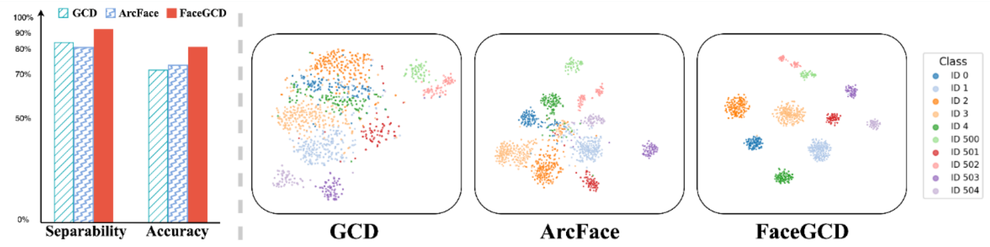
Overview

Challenge: **Generalized Face Discovery (GFD)**
● **Goal**: Simultaneously **identify known faces and discover new ones** from mixed **labeled/unlabeled data**.
● **Difficulty**: **Fails existing GCD and Face Recognition methods** due to the **massive number** of identities and subtle, **fine-grained differences** between them.

Our Approach: **FaceGCD**
We propose **FaceGCD**, which adapts prefixes to frozen Vision Transformer (ViT).
A lightweight **HyperNetwork** generates **instance-specific prefixes** on the fly.
These prefixes are injected into the ViT layers, enabling it to **capture subtle identity cues without expensive retraining**.

Contributions
● **Task**: Introduced **GFD**, **a practical new open-world face recognition task**.
● **Method**: Proposed **FaceGCD**, an **efficient**, **adaptive model** using **dynamic prefixes**.
● **Performance**: Achieved **State-of-the-Art results**, significantly outperforming baselines on large-scale benchmarks.

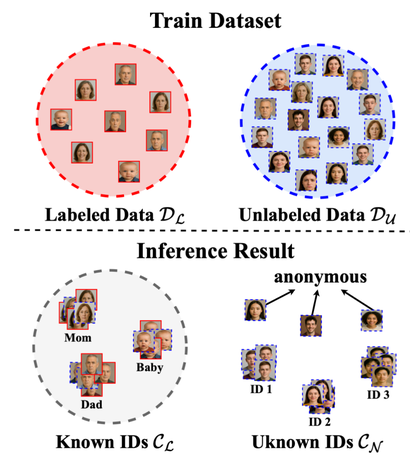
Limitations of Traditional Paradigms for GFD



Limitations of Existing Setups
Existing methods **cannot handle** the challenges of GFD : high cardinality and the need for open-set discovery.

1. Traditional Face Recognition (FR)
● **Focus**: Closed-set identification (only known IDs).
● **Limitations**: Cannot discover novel IDs, resulting in misclassification of new faces.

2. Generalized Category Discovery (GCD)
● **Focus**: Low-cardinality tasks (hundreds of objects) with clear visual differences.
● **Limitations**: Lack of sensitivity to the fine-grained, subtle differences between thousands of visually similar faces, leading to poor clustering of new IDs.



GFD Setup
● **Labeled Set** ($\mathcal{D}_{\mathcal{L}}$) :A set with labels for **some known IDs** ($\mathcal{C}_{\mathcal{L}}$).
● **Unlabeled Set** ($\mathcal{D}_{\mathcal{U}}$) :A set containing **both known** ($\mathcal{C}_{\mathcal{L}}$) and **unknown** ($\mathcal{C}_{\mathcal{N}}$) **IDs**.

Our Method: FaceGCD for Instance-Specific Features

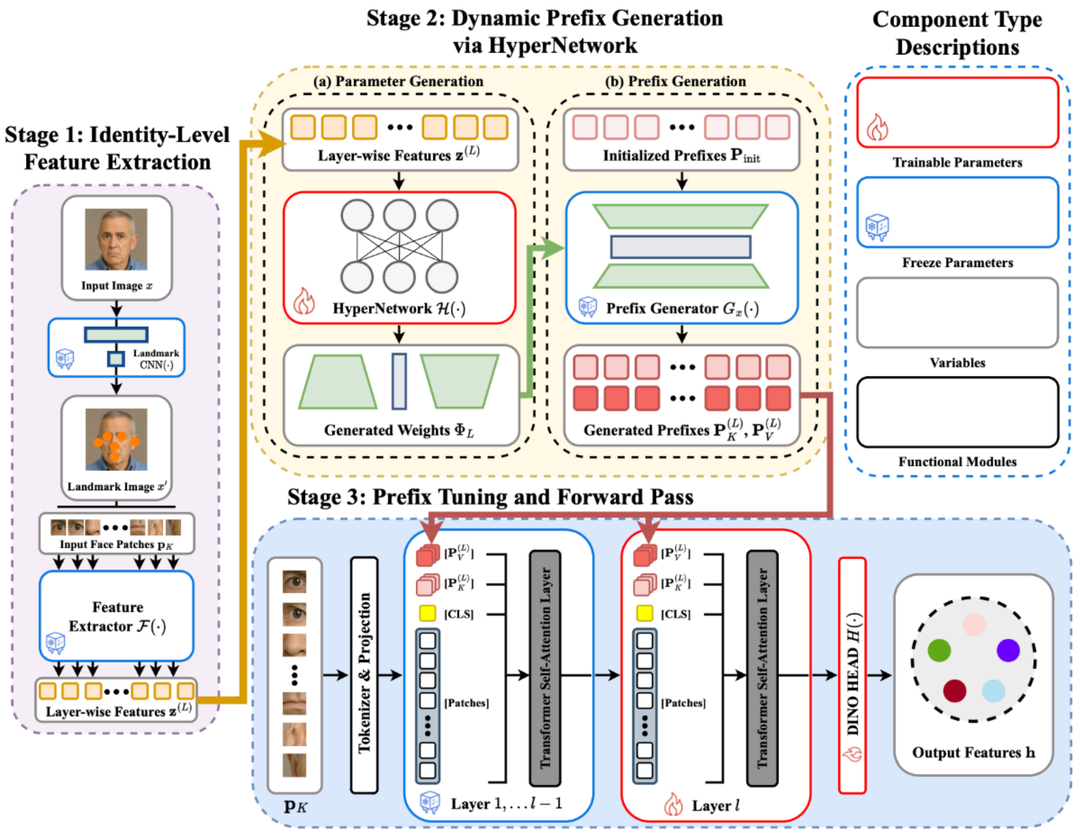
Core Concept
FaceGCD dynamically constructs **an instance-specific feature extractor** for **every input face**. This is achieved by modulating a **powerful**, frozen backbone network with **lightweight**, **dynamically generated prefixes**.

Stage 1: Identity-Level Feature Extraction
● **Component:** Landmark CNN$(\cdot)$, Frozen DINO ViT $\mathcal{F}(\cdot)$
● **Role:** To **provide a robust and stable foundation** of general facial features. The layer-wise Features $\mathbf{z}^{(L)}$, serve as a **compact representation** of the input face.
● **Key Advantage:** The frozen backbone ensures **stability and high parameter efficiency**, as only the **lightweight HyperNetwork** is trained. Crucially, it retains **common and balanced facial knowledge** across all IDs.

Stage 2: Dynamic Prefix Generation via HyperNetwork
(a) Parameter Generation
● **Component**: A lightweight HyperNetwork $\mathcal{H}(\cdot)$ (2-layer MLP)
● Role: To generate a set of **layer specific weights** Φ_L for **Prefix Generator** $G_x(\cdot)$
● **Key Advantage:** Enables **instance-specific adaptation**. The model learns how to generate parameters for a given IDs, **rather than learning static parameters** for all faces.
(b) Prefix Generation
● **Component:** Layer-specific Prefix Generator $G_x(\cdot)$, Prefixes $\mathbf{P}_K^{(L)}, \mathbf{P}_V^{(L)}$
● Role: Each generator configured by its **layer-specific weights** Φ_L, transforms the **random initialized Prefixes** $\mathbf{P}_{\text{init}}$ into **instance-specific Prefixes** $\mathbf{P}_K^{(L)}, \mathbf{P}_V^{(L)}$
● **Key Advantage:** This two-step process is **highly parameter-efficient**. The resulting **instance-specific prefixes** are crucial for **capturing the fine-grained facial cues** that define each unique identity.

$$\mathbf{P}_K^{(L)}(x), \mathbf{P}_V^{(L)}(x) = G_x(\mathbf{P}_{\text{init}}; \mathcal{H}(\mathbf{z}^{(L)}))$$

Stage 3: Prefix Tuning and Forward Pass
● **Component:** Frozen DINO ViT $\mathcal{F}(\cdot)$ (Only last layer trainable), DINO head $H(\cdot)$
● **Role:** The **dynamic prefixes** are prepended to the **Key and Value vectors**, thereby tuning the feature extraction process of the **DINO backbone** specifically for the input image.
● **Key Advantage:** Allows the model to adaptively focus on and **amplify subtle visual cues**, enabling the backbone to capture **instance-specific features** crucial for discriminating among **fine-grained** and **visually similar face IDs**.



$$\tilde{K}^{(L)} = [\mathbf{P}_K^{(L)}; K^{(L)}], \quad \tilde{V}^{(L)} = [\mathbf{P}_V^{(L)}; V^{(L)}], \quad \tilde{Q}^{(L)} = Q^{(L)}$$

: Prepend Prefixes to Key and Value vectors

$$\mathcal{L}_B = (1-\tau)\sum_{i \in B} \mathcal{L}_i^u + \lambda \sum_{i \in B \cap \mathcal{D}_{\mathcal{L}}} \mathcal{L}_i^s$$: Semi-contrastive loss

: Loss Function (Same as GCD baseline)

Result Tables

Method	YTF 500			YTF 1000			YTF 2000			CASIA 500			CASIA 1000			CASIA 2000		
	All	Known	Novel	All	Known	Novel	All	Known	Novel	All	Known	Novel	All	Known	Novel	All	Known	Novel
GCD	70.5	86.4	54.5	70.4	78.1	62.7	70.3	79.9	60.7	40.8	41.0	40.6	45.4	44.2	46.5	47.7	47.2	48.3
SimGCD	74.1	**96.4**	51.8	72.8	**95.1**	50.1	72.6	**92.7**	52.5	31.2	32.3	30.1	39.2	37.7	40.7	41.8	39.4	44.2
PromptCAL	75.2	88.6	61.8	75.0	86.1	63.8	66.6	76.9	56.2	44.9	43.8	45.9	49.9	49.3	50.4	40.3	39.6	41.0
CMS	52.2	73.4	30.9	44.5	61.6	27.4	26.6	32.9	20.2	25.3	26.5	24.1	24.4	25.0	23.7	22.7	23.1	22.3
FaceGCD (Ours)	**81.2**	93.8	**68.6**	**82.7**	93.2	**72.1**	**83.6**	91.8	**75.4**	**58.8**	**68.9**	**48.7**	**52.2**	**52.3**	**52.2**	**56.1**	**60.4**	**51.7**

Table 1. FaceGCD vs existing GCD methods on various GFD benchmarks.

Feature Extractor	K-Means			DBSCAN			HAC			L-GCN			Ada-NETS			SSK		
	All	Known	Novel	All	Known	Novel	All	Known	Novel	All	Known	Novel	All	Known	Novel	All	Known	Novel
ArcFace	66.5	65.4	67.6	13.9	13.9	13.9	68.5	68.7	68.4	33.2	17.7	48.7	24.6	27.2	21.9	73.0	78.3	67.6
ArcFace + GCD	69.6	83.9	55.3	27.5	41.0	14.0	70.1	82.8	57.3	44.6	47.2	41.9	24.9	27.7	22.0	72.0	87.9	56.1
FaceGCD (Ours)	**78.6**	**86.6**	**70.5**	**42.0**	**49.1**	**34.8**	**82.0**	**88.1**	**75.9**	**52.3**	**47.5**	**57.1**	**32.6**	**35.9**	**29.4**	**82.7**	**93.2**	**72.1**

Table 2. FaceGCD vs ArcFace on YTF 1000 with various clustering schemes.

Method	CIFAR100			ImageNet100			CUB			Stanford Cars			FGVC Aircraft			Herbarium 19		
	All	Old	Novel	All	Old	Novel	All	Old	Novel	All	Old	Novel	All	Old	Novel	All	Old	Novel
GCD	73.0	76.2	66.5	74.1	89.8	66.3	51.3	56.6	48.7	39.0	57.6	29.9	45.0	41.1	46.9	35.4	51.0	27.0
SimGCD	80.1	81.2	**77.8**	83.0	93.1	77.9	50.3	65.6	57.7	53.8	71.9	45.0	54.2	59.1	51.8	**44.0**	**58.0**	36.4
PromptCAL	81.2	84.2	75.3	83.1	92.7	78.3	62.9	64.4	62.1	50.2	70.1	40.6	52.2	52.2	**52.3**	37.0	52.0	28.9
CMS	82.3	85.7	75.5	**84.7**	95.6	**79.2**	**68.2**	**76.5**	**64.0**	56.9	**76.1**	47.6	**56.0**	63.4	**52.3**	36.4	54.9	26.4
Ours	**83.4**	**88.8**	72.5	78.4	**95.7**	69.6	64.5	73.1	60.1	**62.5**	52.1	**67.6**	55.4	**63.5**	51.3	42.6	44.6	**40.8**

Table 3. Comparison of generic GCD benchmarks with GCD methods.

Method	All	Known	Novel	Additional Params	Trainable Params	Total Params
FaceGCD (Prefix size: 10)	**82.9**	92.9	**72.8**			
Prefix size: 5	80.9 (-2.1)	91.1 (-1.8)	70.4 (-2.4)	13.8M (6.6%)	41.1M (19.8%)	207.7M
Prefix size: 20	82.7 (-0.2)	**93.2 (+0.3)**	72.1 (-0.7)			
Static Prefix Generator	76.9 (-6.0)	81.7 (-11.2)	72.1 (-0.7)	399.9M (67.3%)	427.2M (72.0%)	593.8M
Static Prefix Pool	42.6 (-40.3)	52.8 (-40.1)	32.3 (-40.5)	**3.5M (1.8%)**	**30.8M (15.6%)**	197.3M

Table 4. Ablation Study on prefix size and alternative prefixing strategies on YTF 1000.

Analysis and Ablation Study

Experimental Results
● **Dominant Performance**: **FaceGCD** outperforms both **FR (ArcFace) and existing GCD methods** in the challenging **GFD scenario**. The **performance gap widens** as the number of **identities increases**, demonstrating the **effectiveness of dynamic prefixes** in handling high-cardinality, fine-grained tasks.
● **Excellent Generality on GCD Benchmarks**: Beyond faces, FaceGCD achieves **competitive SOTA performance** on standard **GCD benchmarks** (e.g., CUB, Stanford Cars), proving the **versatility** of the **instance-specific prefix mechanism.**

Ablation Study: Validating Core Components
● **Dynamic Prefixes are Crucial**: Removing the HyperNetwork entirely and replacing it with static prefixes results in a severe performance degradation. This directly confirms our hypothesis that adaptive, **per-image feature modulation is essential.**
● **HyperNetwork is Effective & Efficient**: The **HyperNetwork** proves to be a **highly effective** and **parameter-efficient module** for generating these dynamic prefixes. It successfully provides the necessary i**nstance-specific conditioning** while adding **minimal computational overhead**, demonstrating its suitability for large-scale applications.

Conclusion
We introduced **Generalized Face Discovery (GFD)**, a **novel** and **practical task** for **open-world face recognition**. Our proposed **FaceGCD**, powered by **dynamic prefix generation**, effectively solves the core GFD challenges of **high cardinality** and **fine-grained similarity**. By establishing a **new state-of-the-art**, **FaceGCD** provides a **strong baseline** for future research in **large-scale, open-world recognition**.