# LoRACLR: Contrastive Adaptation for Customization of Diffusion Models

Enis Simsar[1] Thomas Hofmann[1] Federico Tombari[2,3] Pinar Yanardag[4]

# Introduction

- **Objective:** Seamlessly merge multiple LoRA models (one per concept) into a unified diffusion model.
- **Significance:** Existing methods struggle with identity entanglement or require joint retraining.

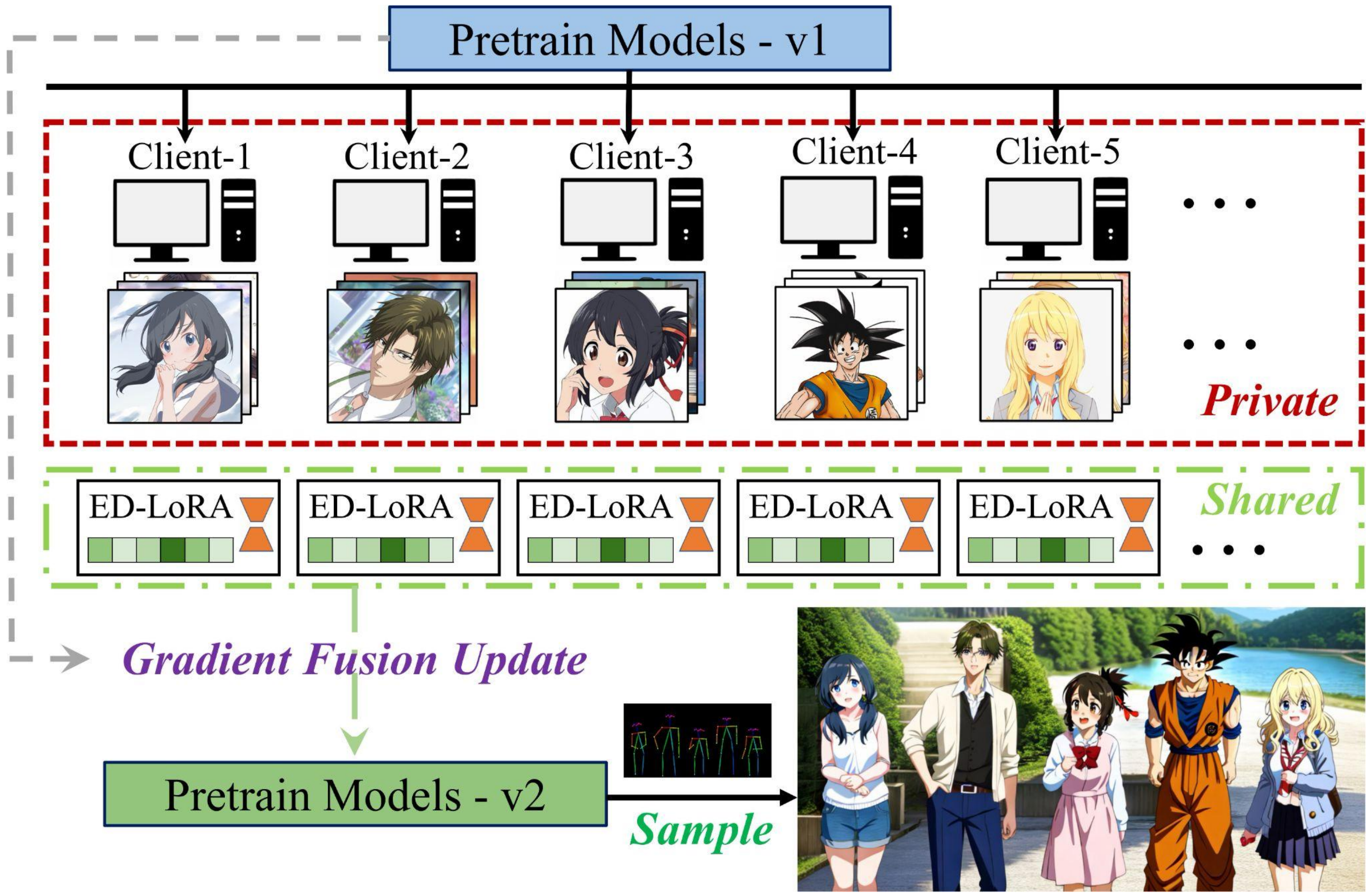
 &  &  , in a castle, signing papers, in  style...

# Contributions

- **Contrastive Merging Objective**
  - Aligns concept-specific LoRA outputs while preserving identity.
  - Repels unrelated features to reduce interference.
- **No Need for Retraining**
  - Operates post-hoc using pre-trained LoRA models.
- **Scalability & Efficiency**
  - Merges up to 12 LoRAs in ~5 minutes.
  - Output quality holds up even with many concepts.

# Mix-of-Show[1]

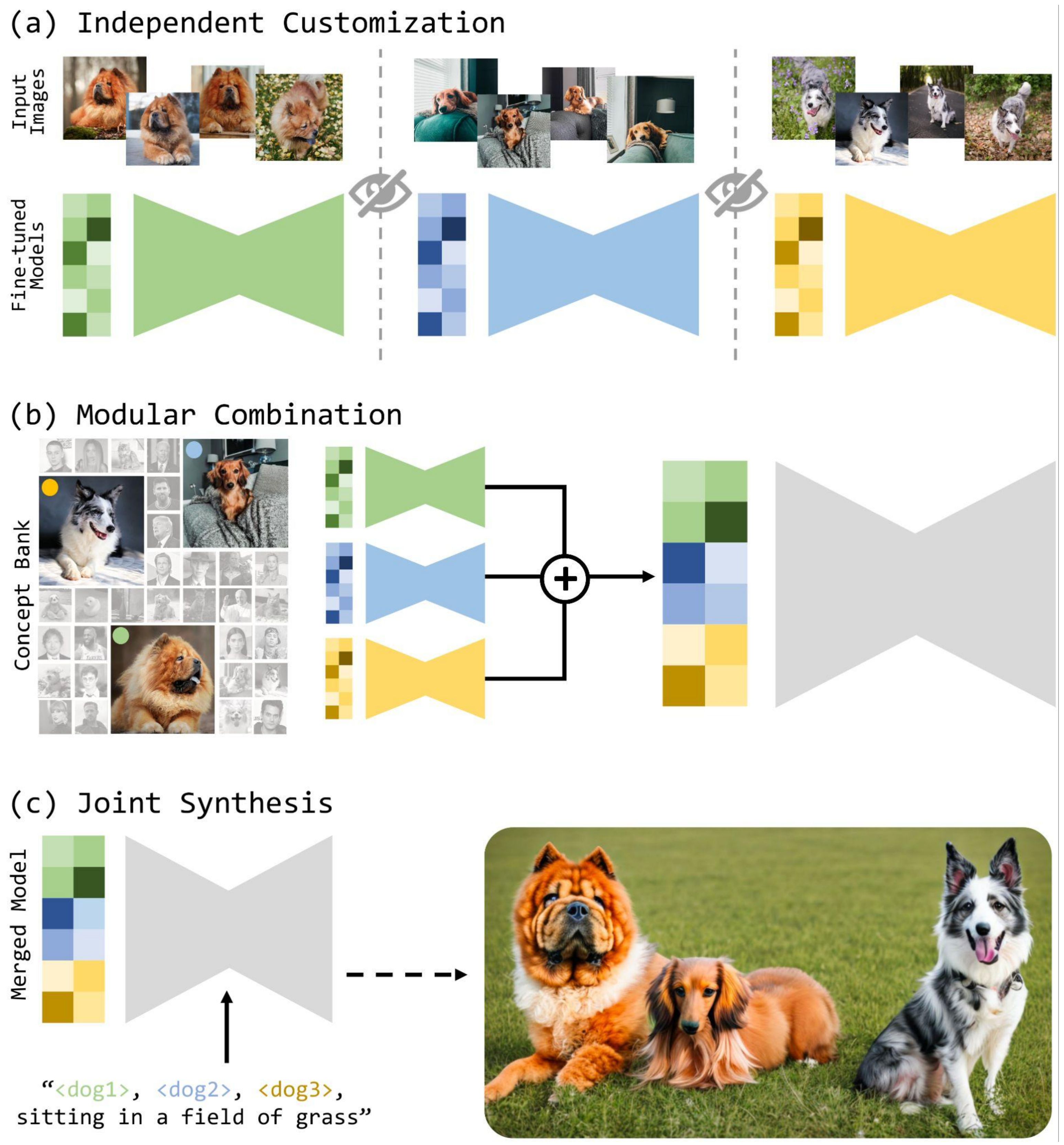
- Merges multiple concepts using Embedding-Decomposed LoRAs (ED-LoRA)
- Requires access to original training data
- Proposes decentralized training of LoRAs per concept → Each LoRA trained independently, merged via gradient fusion
- Incompatible with community LoRAs




[1] Gu, Yuchao, et al. "Mix-of-show: Decentralized low-rank adaptation for multi-concept customization of diffusion models." In NeurIPS. 2023.

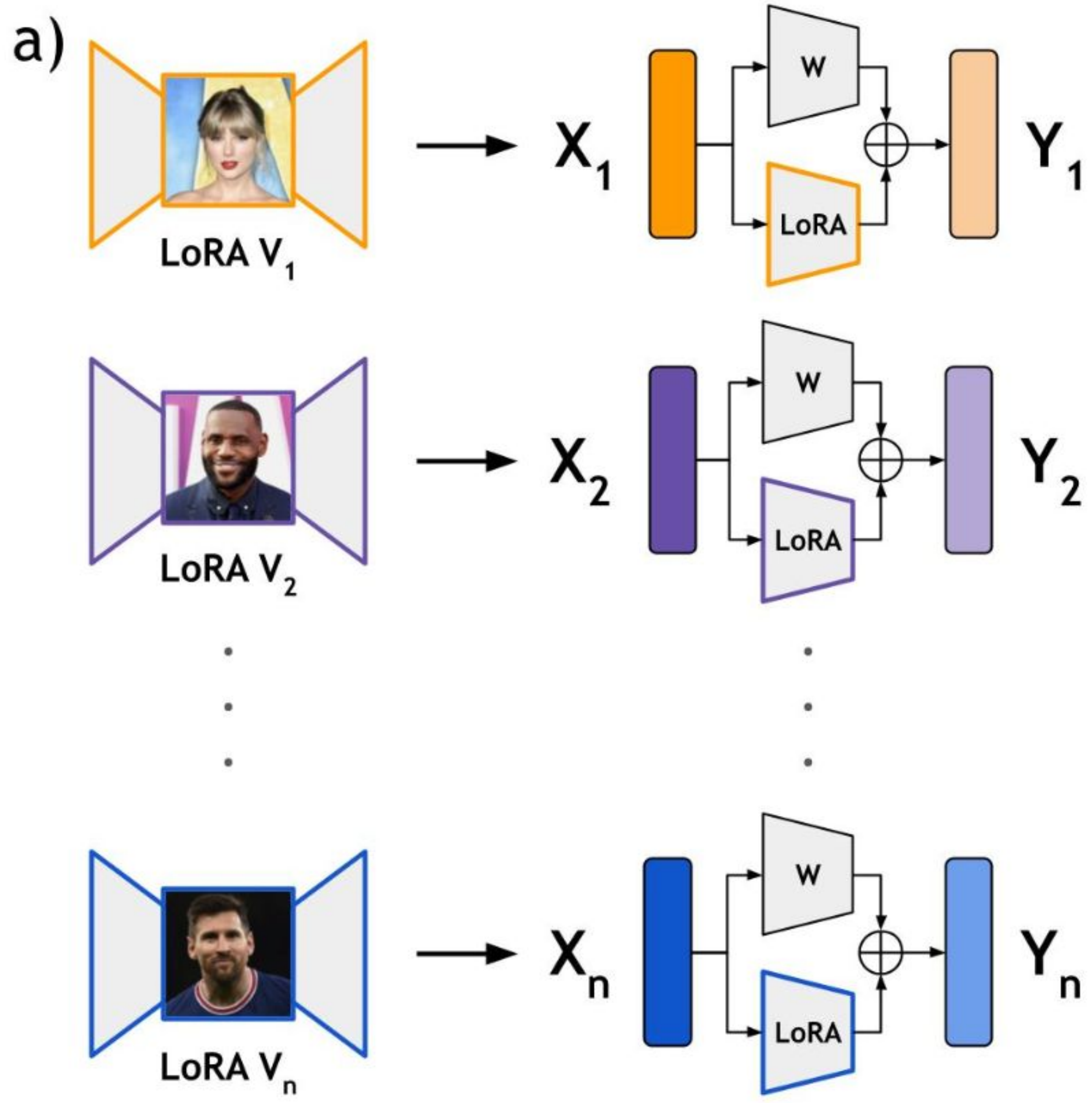
# Orthogonal Adaptation[2]

- Enforces orthogonality across LoRA directions during fine-tuning
- Mitigates interference, but:
  - Requires access to original training data
  - Requires retraining from scratch
  - Less efficient for large-scale merging since it requires training data and retraining




[2] Po, Ryan, et al. "Orthogonal adaptation for modular customization of diffusion models." In CVPR. 2024.

# Collecting Features from Individual LoRAs

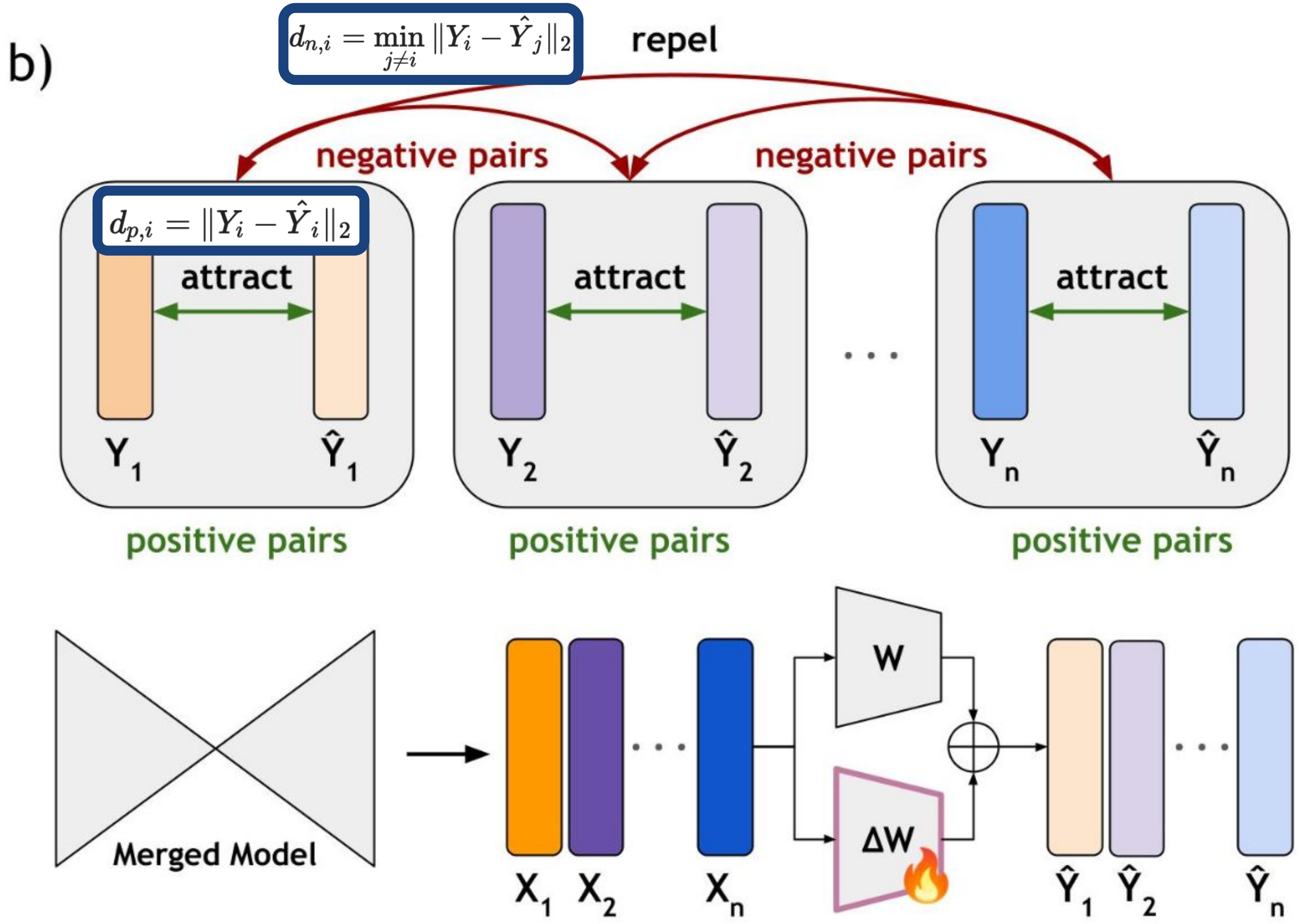


# Merging LoRAs by applying Contrastive Objective

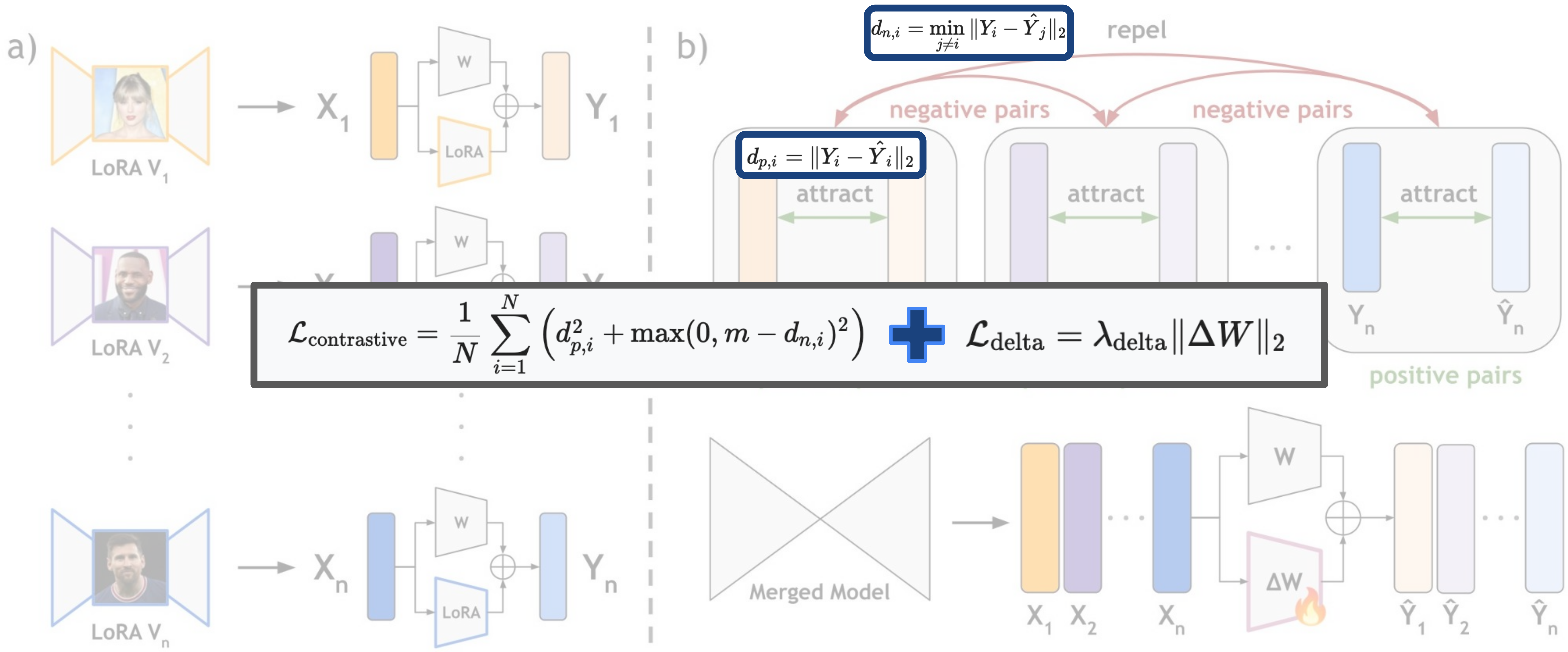
- **Delta-Based Merging**

Learns a delta ΔW added to base weights, not touching original weights.

- **Contrastive Loss**
  - *Positive pairs:* original LoRA vs. merged output for same concept → attract.
  - *Negative pairs:* original vs. merged for different concepts → repel.



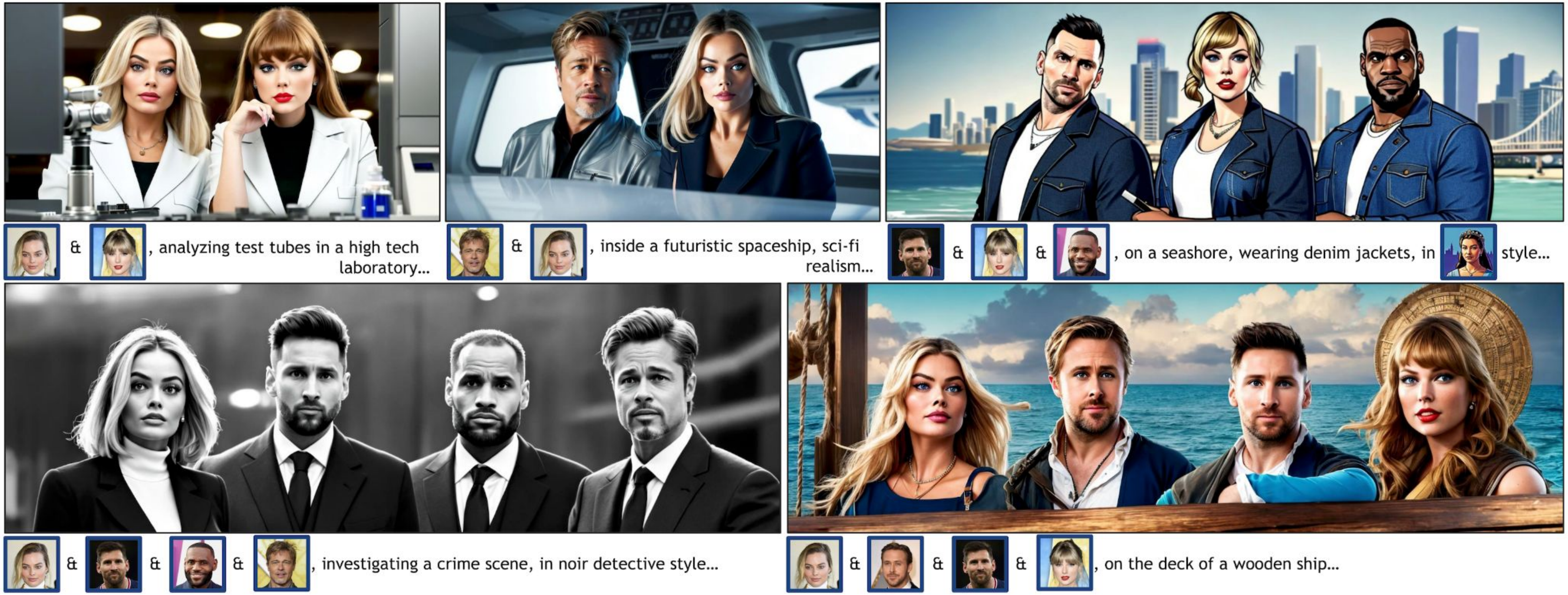
# Overall Objective




$$\mathcal{L}_{\text{contrastive}} = \frac{1}{N}\sum_{i=1}^{N}\left(d_{p,i}^2 + \max(0, m - d_{n,i})^2\right) + \mathcal{L}_{\text{delta}} = \lambda_{\text{delta}}\|\Delta W\|_2$$

# Qualitative Results

- Preserves visual identity across subjects.
- Works across diverse styles and scenes (comic, sci-fi, oil painting).



# Qualitative Comparison

- LoRACLR preserves individual identities, even with six distinct subjects.
- Competing methods show identity drift and visual inconsistency.

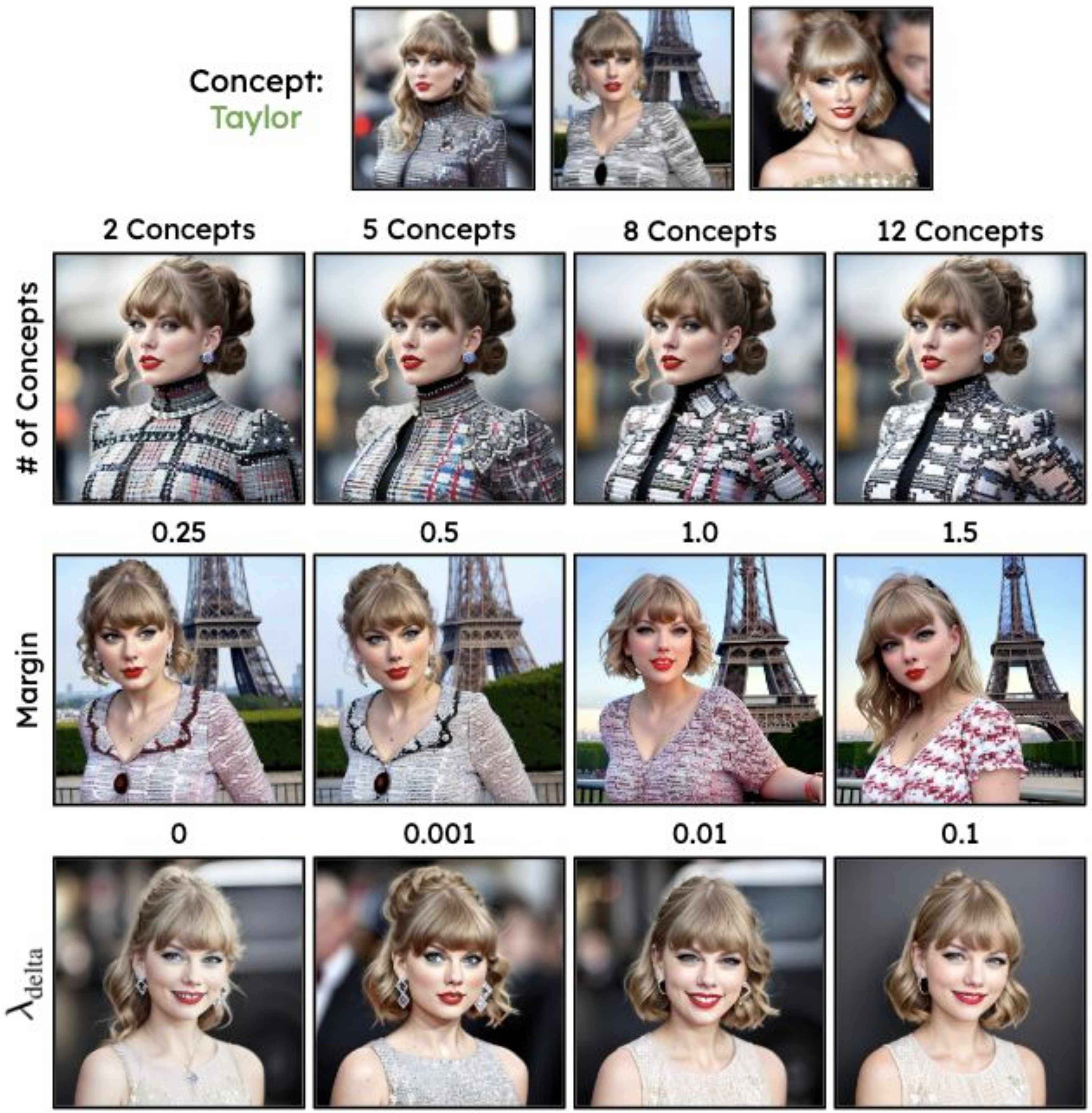
# Quantitative Results

| Methods | Text Alignment ↑ | | | Image Alignment ↑ | | | Identity Alignment ↑ | | |
|---|---|---|---|---|---|---|---|---|---|
| | Single | Merged | Δ | Single | Merged | Δ | Single | Merged | Δ |
| Prompt+ | .643 | .643 | — | .683 | .683 | — | .515 | .515 | — |
| Custom Diffusion | .668 | .673 | +.005 | .648 | .623 | -.025 | .504 | .408 | -.096 |
| DB-LoRA | .613 | .682 | **+.069** | .744 | .531 | -.213 | .683 | .098 | -.585 |
| Mix-of-Show (FedAvg) | .625 | .621 | -.004 | .745 | .735 | -.010 | .728 | .706 | -.022 |
| Mix-of-Show (GradFusion) | .625 | .631 | +.006 | .745 | .729 | -.016 | .728 | .717 | -.011 |
| Orthogonal Adaptation | .624 | .644 | <u>+.020</u> | .748 | .741 | <u>-.007</u> | .740 | .745 | <u>+.005</u> |
| LoRACLR (Ours) | .668 | .665 | -.003 | .766 | .776 | **+.010** | .799 | .828 | **+.029** |

**Bold** indicates the best while <u>Underline</u> indicates the second best.

# Ablation Study

- **Margin variation:** Shows over- or under-separation of identities
- **$\lambda_{delta}$ variation:** Low λ keeps fidelity, high λ over-regularizes
- **Scaling:** Performance holds as concept count increases



# Complex Scenes & Prompts



 & , holding hands walking through a field of tall grass...

 is wearing clothes in noir detective style,

 has the crown on the top of head, and they are on the deck of a wooden ship.

# Conclusion

- LoRACLR introduces a novel post-training approach for merging independently trained LoRA models into a unified diffusion model, enabling multi-concept image synthesis with minimal interference.
- Through a contrastive learning objective, LoRACLR maintains the distinctiveness of each concept while allowing them to coexist in rich, prompt-driven scenes.
- LoRACLR unlocks new potential for creative applications like personalized art, storytelling, and design while also raising important questions about responsible use of generative AI technologies.

# Thank you!

ExHall D Poster #242

Sat 14 Jun 10:30 a.m. CDT — 12:30 p.m. CDT