Multi-Task Learning for Image Inpainting and Matting

Modern computervision increasingly demands sophisticatedpixel-level understanding,particularly forimage manipulation tasksrequiringfine-graineddetailpreservation. This researchproposal explores the synergistic potential of combining image inpainting and image matting within a unified multi-task learning framework, leveraging shared representations to enhance both reconstruction quality and boundary estimation accuracy.

Image Inpainting

Reconstructs missing or corrupted image regions using local texture context and global semantic understanding. Challenges include maintaining structural consistency and realistic detail synthesis at object boundaries.

Image Matting

Estimates per-pixel alpha transparency for foreground-background separation. Requires precise boundary localization and handling of semi-transparent regions like hair or fur textures.

Architectural Framework

1

Shared Encoder

Transformer-based vision encoder captures semantic features and textural patterns common to both tasks

2

Task-SpecificDecoders

Separate heads for pixel synthesis (inpainting) and alpha matte prediction (matting)

3

Cross-TaskConsistency

Alignment losses encourage coherent boundary handling between tasks

Loss Functions and Evaluation

Inpainting Objectives

- L1/L2 pixel reconstruction loss
- Perceptual loss for feature consistency
- Adversarial loss for photorealistic synthesis

Metrics: PSNR, SSIM, FID for quality assessment

Matting Objectives

- Alpha matte regression loss
- Compositional reconstruction loss
- Boundary refinement constraints

Metrics: SAD, MSE, gradient error for boundary precision