OR Paper Review
MaGIC: Multi-modality Guided Image Completion

Bruno Sánchez Gómez

April 20, 2025

Table of Contents

Multi-modal Guided Image Completion (MaGIC)

- **Problem:** Vanilla image completion struggles with large missing regions; existing guided methods often use only a single modality.
- **MaGIC Solution:** A flexible framework for image completion guided by single or *arbitrary combinations* of modalities (Text, Edge, Sketch, Segmentation, Depth, Pose).
- **Components:**
  - **Modality-specific Conditional U-Net (MCU-Net):** Injects single-modal guidance into a U-Net denoiser.
  - **Consistent Modality Blending (CMB):** Training-free method to blend guidance from multiple pre-trained MCU-Nets via latent space gradients. Enables easy addition of new modalities.
- **Results:** Outperforms SOTA methods and generalizes well to various completion tasks.

Examples

Image examples from the paper (maybe more than 1 slide)

MCU-Net

Explain MCU-Net

CMB

Explain CMB

Why did MaGIC succeed?

Where did MaGIC fail?

Future Implications [Placeholder]

- **Generalization:** MaGIC's framework can be applied to other image generation tasks, such as inpainting or super-resolution.
- **Modality Fusion:** The CMB method can be extended to fuse more complex modalities, such as audio or video.
- **Real-world Applications:** Potential applications in fields like medical imaging, autonomous driving, and augmented reality.

References

- **MaGIC: Multi-modality Guided Image Completion**
  https://arxiv.org/abs/2303.14100