OR Paper Review
MaGIC: Multi-modality Guided Image Completion

Bruno Sánchez Gómez

April 20, 2025

Table of Contents

Image Completion

Definition

Image completion refers to the task of filling in missing regions within an image in a visually plausible way.

- **Applications:**
  - **Inpainting:** Restoring damaged or missing parts of an image.
  - **Outpainting:** Extending the boundaries of an image.
  - **Editing:** Modifying images by adding or removing elements.

Approaches to Image Completion

- **Vanilla Image Completion:**
  - Relies solely on existing image pixels around the masked region.
  - Struggles with large missing areas due to limited internal context.
  - Often leads to blurry or repetitive textures.
- **Guided Image Completion:**
  - Uses external cues (e.g., text descriptions, edge maps, segmentation masks) for guidance.
  - Improves results significantly, especially for large gaps, by providing external semantic information.
  - Existing methods often restricted to *single-modality* guidance, limiting flexibility and performance in complex scenarios requiring multiple constraints.

Multi-modal Guided Image Completion (MaGIC)

- **MaGIC:** A flexible framework for image completion guided by single or *arbitrary combinations* of modalities, such as:
  - Text
  - Canny Edge
  - Sketch
  - Segmentation
  - Depth
  - Pose
- **Results:** Outperforms SOTA methods and generalizes well to various completion tasks.

Table of Contents

Examples

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

Components of MaGIC

- **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.

MCU-Net

Explain MCU-Net

CMB

Explain CMB

Table of Contents

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