ETH zürich

MTC
media technology center

Fuzzy-Conditioned Diffusion and Diffusion Projection Attention Applied to Facial Image Correction

Majed El Helou

IEEE ICIP 2023 – Kuala Lumpur

11 - October - 2023

Image inpainting meets image restoration

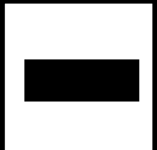


Image inpainting:

Masked image
+
Binary mask area
⬇
Completed image

Required

Prior + Signal

Image restoration:

Degraded image
+
Degradation model info
⬇
Restored image

Contributions:

1. Propose a generalization of inpainting to non-binary cases
2. Derive a non-binary conditional diffusion solution for #1
3. Cast the image restoration problem into fuzzy inpainting
4. Exploit diffusion space statistics to auto-detect anomalies
5. Combine #4 with our conditional diffusion to restore images

vs. RePaint [17]

vs. DDPM [2] projection

Mathematical formulation

Markovian noising process:

$$q(x_t|x_{t-1}) := \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t\mathbb{1})$$

Direct latent sampling:

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)\mathbb{1})$$

$$\begin{cases} \bar{\alpha}_t := \prod_{d=0}^{t} \alpha_d \\ \alpha_t := 1 - \beta_t \end{cases}$$

Bayes' posterior distribution:

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \tilde{\mu}(x_t, x_0), \tilde{\beta}_t\mathbb{1})$$

Conditional posterior distribution parameters:

$$\begin{cases} \tilde{\mu}(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t \\ \tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \end{cases}$$

Network-learned distribution:

$$p_\theta(x_{t-1}|x_t) := \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$$

Residual learning target:

$$E_{t, x_0 \sim q(x_0), \epsilon \sim \mathcal{N}(0, \mathbb{1})}[||\epsilon - \epsilon_\theta(x_t, t)||^2]$$

(with fixed covariance)

Conditional diffusion and distribution equalization

Repeat

Conditional inpainting diffusion:

1. Diffuse a hallucinated image step by step
2. Project the context image into the diffusion
3. Spatially fuse the two intermediate images ← Distribution clash
4. Project backwards to expand spatial information

$$x_m := \sqrt{\bar{\alpha}_{t-1}}x + \frac{m \cdot x^r_{t-1} + (1-m)\cdot x_{t-1} - \sqrt{\bar{\alpha}_{t-1}}x}{\sqrt{1 - 2\cdot m + 2\cdot m^2}}$$

$$m^2 V\{x^r_{t-1}\} + (1 - m)^2 V\{x_{t-1}\}$$

← Fused-image variance

Algorithm pseudocode overview

Algorithm 1 Our fuzzy-conditioned image diffusion

$x_T \sim \mathcal{N}(0, \mathbb{1})$
for $t = T, \dots, 1$ **do**
  for $j = 1, \dots, J$ **do**
    $\epsilon \sim \mathcal{N}(0, \mathbb{1})$ if $t > 1$ else $\epsilon = 0$ ← #1 Sample & diffuse hallucination
    $x^r_{t-1} = \sqrt{\bar{\alpha}_{t-1}} x + \epsilon \sqrt{1 - \bar{\alpha}_{t-1}}$
    $\epsilon_2 \sim \mathcal{N}(0, \mathbb{1})$ if $t > 1$ else $\epsilon_2 = 0$ ← #2 Project the context image
    $x_{t-1} = \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}})\epsilon_\theta(x_t, t) + \tilde{\beta}_t * \epsilon_2$
    $x_m = m \cdot x^r_{t-1} + (1 - m) \cdot x_{t-1} - \sqrt{\bar{\alpha}_{t-1}} x$ ← #3 Fuse & correct the variance
    $x_m = \sqrt{\bar{\alpha}_{t-1}} x + x_m / \sqrt{1 - 2 \cdot m + 2 \cdot m^2}$
    if $j < J$ and $t > 1$ **then**
      $x_t \sim \mathcal{N}(\sqrt{1 - \beta_t} x_{t-1}, \beta_t \mathbb{1})$ ← #4 Backwards projection steps
    end if
  end for
end for
return x_0

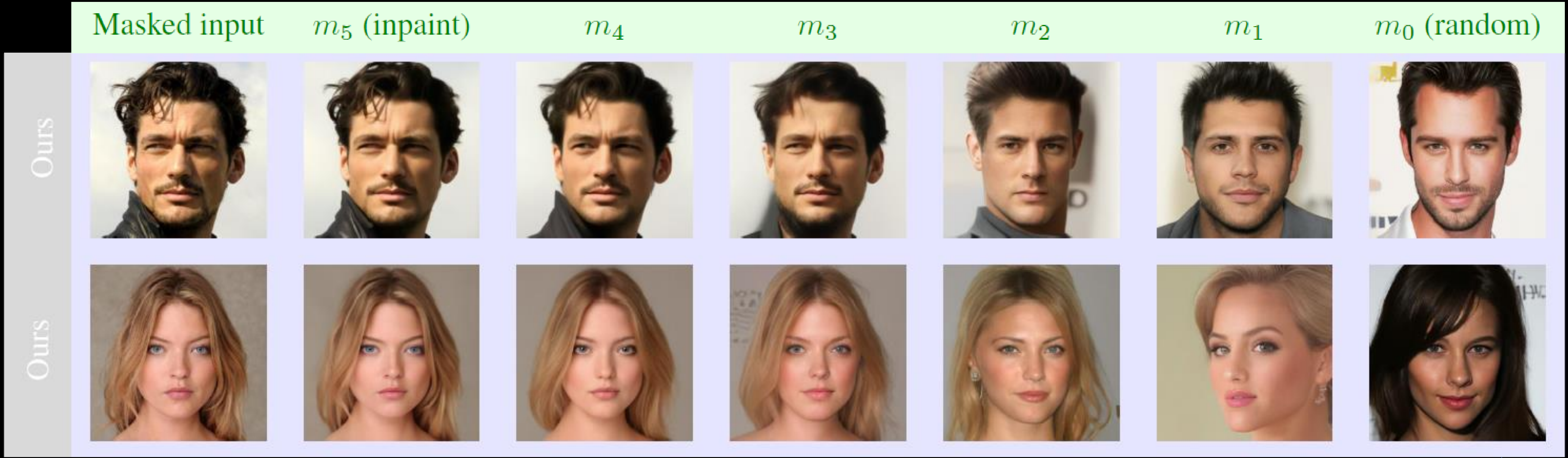
Fuzzy inpainting enables out-of-mask flexibility

$$x_m := \sqrt{\bar{\alpha}_{t-1}}x + \frac{m \cdot x^r_{t-1} + (1-m) \cdot x_{t-1} - \sqrt{\bar{\alpha}_{t-1}}x}{\sqrt{1 - 2 \cdot m + 2 \cdot m^2}}$$

Decreasing the value of the conditioning weight map (that multiplies the inpainting mask)

Fuzzy-conditioned controllable generation

$$x_m := \sqrt{\bar{\alpha}_{t-1}}x + \frac{m \cdot x^r_{t-1} + (1-m)\cdot x_{t-1} - \sqrt{\bar{\alpha}_{t-1}}x}{\sqrt{1 - 2\cdot m + 2\cdot m^2}}$$



Decreasing the value of the conditioning weight map

Diffusion projection attention map

Objective: Degradation-agnostic and autonomous anomaly detection
Approach: Compare diffusion projection discrepancy to normal statistics

$$A(x) := \frac{1}{N} \sum_{t \in PS} \mathcal{H}\left(\frac{\phi(x - \widehat{x}_t) - \mu_t\{\mathcal{V}\}}{\sigma_t\{\mathcal{V}\}} \right)$$

Min-max to operate between σ and 6σ:

$$\mathcal{H}(x) \;=\; min(max(x, 1), 6)$$

Projecting on central diffusion depths:

$$\{300, 400, 500, 600\}$$

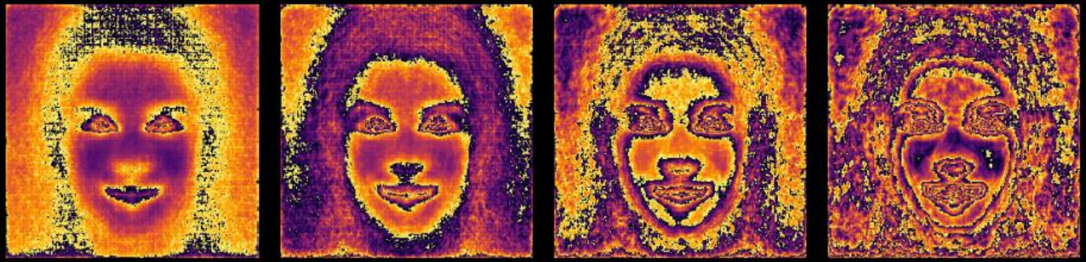
$$\mu_t\{\mathcal{V}\} := \frac{1}{|\mathcal{V}|} \sum_{z \in \mathcal{V}} \phi(z - \widehat{z}_t)$$

$$\sigma_t\{\mathcal{V}\} := \sqrt{\frac{1}{|\mathcal{V}|} \sum_{z \in \mathcal{V}} (\phi(z - \widehat{z}_t) - \mu_t\{\mathcal{V}\})^2}$$

$$m := (1 - A(x))^2$$

Inverted attention corresponds to the hallucination weight
Non-linear scaling going from σ-based distance to weights

Projection space means, and degradation attention

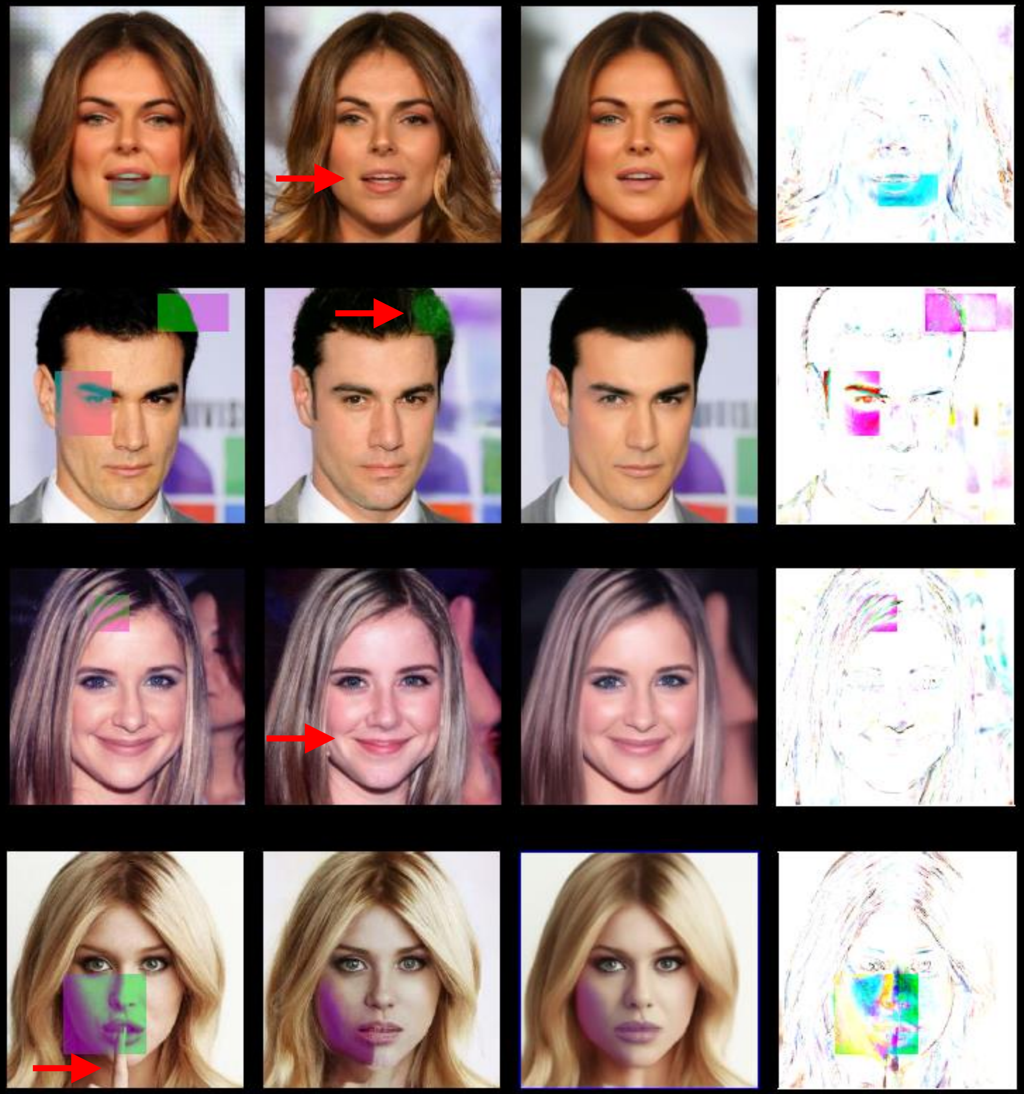


(a) $\mu_{300}\{\mathcal{V}\}$ (b) $\mu_{400}\{\mathcal{V}\}$ (c) $\mu_{500}\{\mathcal{V}\}$ (d) $\mu_{600}\{\mathcal{V}\}$

(e) Input x (f) $\phi(x - \widehat{x}_{300})$ (g) $\phi(x - \widehat{x}_{400})$ (h) $m(x)^*$

(i) Input x (j) $\phi(x - \widehat{x}_{500})$ (k) $\phi(x - \widehat{x}_{600})$ (l) $m(x)^*$

Sample results of diffusion-guided auto-correction



(a) Input x (b) DDPM [2] (c) Ours (d) Our $m(x)^*$

Conclusion limitations and future work

Contributions:
Propose a generalization of inpainting to non-binary cases
Derive a non-binary conditional diffusion solution for #1
Cast the image restoration problem into fuzzy inpainting
Exploit diffusion space statistics to auto-detect anomalies
Combine #4 with our conditional diffusion to restore images

Limitations:
The prior is based on a diffusion model trained on facial images
Unexpected real outlier content risks being classified as anomaly

Future work:
Performing the prior and signal fusion in the frequency domain

ETH zürich

MTC
media technology center

Q & A

Majed El Helou

IEEE ICIP 2023 – Kuala Lumpur

11 - October - 2023