Computational Imaging Project Proposal

Kartikaeya Kumar, Akash Haridas, Zakaria Patel

1 Introduction and Motivation

Deconvolution is the problem of estimating an image from a blurred and possibly noisy measurement. Restoring a blurry image to its true form has applications ranging from everyday photography to medical and scientific imaging. Image deconvolution is often framed as an inverse problem and solved using iterative optimization techniques incorporating priors.

Alternating-Direction Method of Multipliers (ADMM) [1] is a general optimization algorithm for inverse problems. It allows us to incorporate priors based on our understanding of the characteristics of the original image. Moreover, it allows us to mix and match different image formation models (how the image was corrupted) and regularizers/priors (how we think a natural image looks) [2]. In particular, an existing image denoiser can be used as a prior to produce clean images when deblurring under Gaussian noise.

The prior has a significant impact on the performance of ADMM. For example, a denoising CNN (DnCNN) prior outperforms a TV prior. As such, extending to other learned denoising priors may yield further improvements.

Denoising diffusion probabilistic models [3] (referred to simply as diffusion models) are a recent development in deep learning and are currently state-of-the-art in image generation. Diffusion models work by iteratively denoising an initial image to invert a forward noising process. The iterative nature of the diffusion denoising process allows it to be guided or conditioned based on a text prompt or class [4].

In this project, we propose using a text-guided diffusion denoiser as the prior in deconvolution with ADMM.

2 Related Work

2.1 Deblurring with ADMM

For inverse imaging applications such as deconvolution, the observed images can be modeled as,

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \eta \tag{1}$$

where $\mathbf{b}$ is the measured image, $\mathbf{A}$ is the convolution kernel expressed as a circulant Toeplitz matrix, $\mathbf{x}$ is the original image on which the convolution kernel operates and η is zero mean Gaussian noise. More formally, $x_i \sim \mathcal{N}(x_i, 0)$, $\eta_i \sim \mathcal{N}(0, \sigma^2)$, which leads to $b_i \sim \mathcal{N}((\mathbf{A}\mathbf{x})_i, \sigma^2)$. Formulating this as a bayesian maximum-a-posterior (MAP) problem, the optimal solution can be obtained as,

$$\mathbf{x}_{MAP} = arg\, min_x \frac{1}{2\sigma^2}||\mathbf{b} - \mathbf{A}\mathbf{x}||_2^2 - \log p(\mathbf{x}) \tag{2}$$

where the second term, $\Psi(\boldsymbol{x}) = -\log p(\boldsymbol{x})$ encodes a known prior on the image (such as sparsity or smoothness). The Alternating Direction Method of Multipliers (ADMM) [1] is a iterative algorithm for convex optimization which has been shown to outperform other optimization techniques in terms of visual quality on such inverse problems. Often, ADMM approaches exploit a known prior to converge to a more likely solution. For example, we can use the lasso ($L1$ norm) prior to promote sparsity, the total variation prior for sparsity of gradients, non-local means to promote self-similarity, Wiener deconvolution to promote direct inverse filtering, and DnCNN for learning-based approaches. Together, these priors form the baseline for comparison to our proposed method discussed in the next section.

2.2 Denoising diffusion models

Recently, diffusion models have succeeded in generating high-quality images from pure Gaussian noise [3]. The training process consists of two steps: First, a forward process (called diffusion) where noise is gradually added to the image sample over T steps until it resembles pure noise, and second, a reverse process where the original image is recovered by iteratively learning the added noise which needs to be subtracted from the current image to obtain the next less-noisy image in the sequence. The resulting generative model can then be sampled by iteratively denoising either a pure Gaussian noise sample or a partially noisy image.

The forward process is a Markov chain that gradually adds Gaussian noise to the image as,

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) = \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}) \tag{3}$$

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta}\mathbf{x}_{t-1}, \beta\mathbf{I}) \tag{4}$$

The reverse process to obtain the sample image, $\mathbf{x}_0$, from the noisy image, $\mathbf{x}_T$, is modelled as,

$$p_\theta(\mathbf{x}_{0:T}) = p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) \tag{5}$$

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \beta\mathbf{I}) \tag{6}$$

where a U-Net is used to estimate $\mu_\theta(\mathbf{x}_t, t)$, the learned noise in $\mathbf{x}_t$, so that $\mathbf{x}_{t-1}$ can be simply computed as,

$$\mathbf{x}_{t-1} = \mathbf{x}_t - \mu_\theta(\mathbf{x}_t, t) \tag{7}$$

CLIP [5] is a model that provides a semantic similarity measure between image-text pairs. Recent methods like

OpenAI's GLIDE and DALL·E 2 [4], [6] guide the diffusion process such that the CLIP-similarity between a generated image and a provided text prompt is maximized, resulting in an effective text-to-image generator.

3 Project Proposal

We propose using a diffusion denoising model as the prior in ADMM. Similar to how denoised images from the DnCNN are used in the z-update to provide a prior on how clean, natural images should look, we plan to experiment with using the denoised image from the diffusion model as the prior.

An advantage of this method is that the diffusion denoiser can be guided by text prompts to provide a stronger prior on how the image should look. For example, a blurred and noisy image of a dog can be deconvolved by ADMM with a denoising prior guided by the text prompt "Dog". We plan to perform experiments to study the effects of such a procedure. In particular, we aim to test whether this method produces better quality images when the noise level is extremely high.

We plan to compare this technique to several baselines such as inverse and Wiener filtering, and ADMM baselines using the TV, DnCNN, and NLM priors, especially at high noise levels.

3.1 Milestones and Timelines

Our proposed timeline is as follows:

- Nov 17 - 22: Acquire an appropriate dataset, and get baseline performance measures (Wiener filter, ADMM with TV, DnCNN and NLM priors).
- Nov 23 - 28: Implement diffusion denoising and GAN-based denoising priors in ADMM. Evaluate performance of these algorithms compared to baselines.
- Nov 29 - Dec 3: Experiment with text-guided diffusion model as the prior in ADMM.
- Dec 3 - Dec 6: Organize results, write the report, and prepare presentation and poster.

References

[1] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[2] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *2013 IEEE Global Conference on Signal and Information Processing*. IEEE, 2013, pp. 945–948.

[3] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851, 2020.

[4] A. Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. McGrew, I. Sutskever, and M. Chen, "Glide: Towards photorealistic image generation and editing with text-guided diffusion models," *arXiv preprint arXiv:2112.10741*, 2021.

[5] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark *et al.*, "Learning transferable visual models from natural language supervision," in *International Conference on Machine Learning*. PMLR, 2021, pp. 8748–8763.

[6] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with clip latents," *arXiv preprint arXiv:2204.06125*, 2022.