Deep Learning for Computer Vision

Classifier and Classifier-Free Diffusion Guidance

Vineeth N Balasubramanian

Department of Computer Science and Engineering
Indian Institute of Technology, Hyderabad

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Review

Question

- In the previous lecture, we saw diffusion methods that use noise as input to generate images. Can we generate an image for a desired output?

Review

Question

- In the previous lecture, we saw diffusion methods that use noise as input to generate images. Can we generate an image for a desired output? Yes! **See this paper and this paper; they leverage Classifier and Classifier-Free Guidance, which we will talk about now**

Unconditional Generation

Problem

- Diffusion goes from noise to images step-by-step
- Not directly suitable to generate a desired output
- Due to their stochastic (random) nature, the generation cannot be controlled

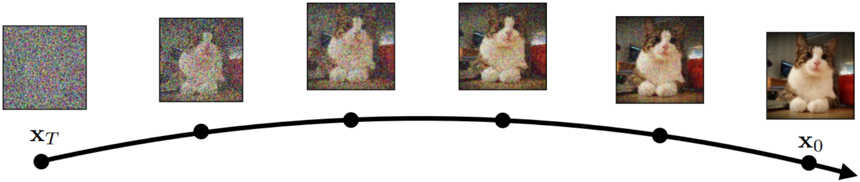
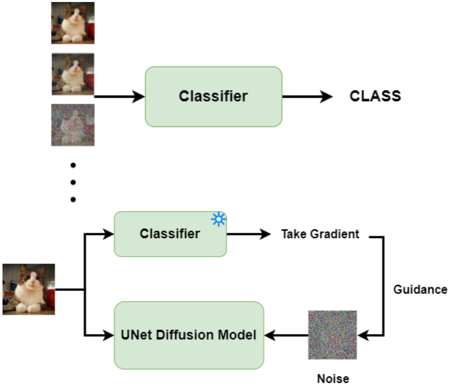


Image Credit: *CVPR 2023 Tutorial on Diffusion Models*

Classifier Guidance Generation

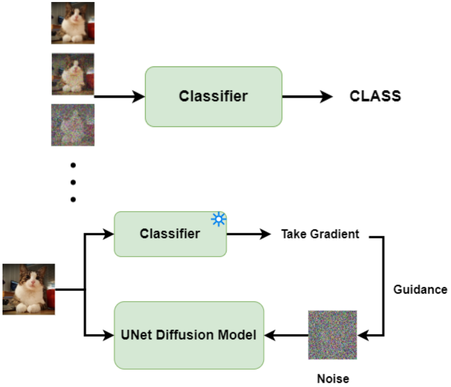
- Perturb the denoising trajectory. How?



Dhariwal et al, Diffusion Models Beat GANs on Image Synthesis, NeurIPS 2021

Classifier Guidance Generation

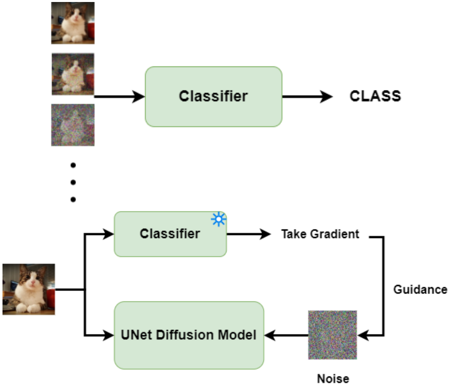
- Perturb the denoising trajectory. How?
- Train a noisy image classifier $p(y|x)$, freeze the classifier and learn the diffusion model



Dhariwal et al, Diffusion Models Beat GANs on Image Synthesis, NeurIPS 2021

Classifier Guidance Generation

- Perturb the denoising trajectory. How?
- Train a noisy image classifier $p(y|x)$, freeze the classifier and learn the diffusion model
- Perturb using gradients of a classifier $\gamma\nabla_{x_t} \log p(y|x_t)$



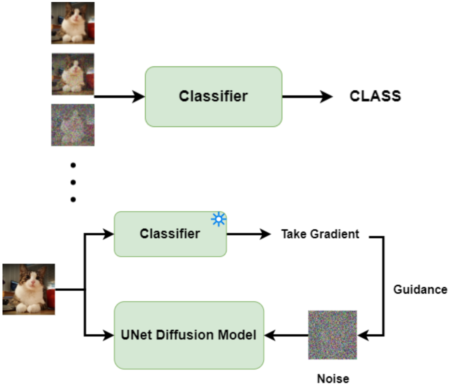
Dhariwal et al, Diffusion Models Beat GANs on Image Synthesis, NeurIPS 2021

Classifier Guidance Generation

- Perturb the denoising trajectory. How?
- Train a noisy image classifier $p(y|x)$, freeze the classifier and learn the diffusion model
- Perturb using gradients of a classifier $\gamma \nabla_{x_t} \log p(y|x_t)$
- γ (should be greater than 1) is used to amplify the influence of the conditioning signal:

$$\widehat{\epsilon}(x_t, t, y) = \epsilon_\theta\,(x_t, t) - \gamma \nabla_{x_t} \log p(y|x_t)$$

- We essentially sample gradients from a classifier when classifying an image of a desired class and feed that gradient information into the diffusion model to lead it to generate the desired class



Dhariwal et al, *Diffusion Models Beat GANs on Image Synthesis*, NeurIPS 2021

Classifier Guidance Generation: How does it work?

- To introduce the class-conditioning, the probability $p(x_t)$ becomes conditional probability $p(x_t|y)$, i.e. $\nabla_{x_t} \log p(x_t) \rightarrow \nabla_{x_t} \log p(x_t|y)$
- Using Bayes Theorem, the new score function for the diffusion model becomes:

$$\nabla_{x_t} \log p(x_t|y) = \nabla \log \left(\frac{p(x_t)p(y|x_t)}{p(y)} \right) = \underbrace{\boxed{\nabla \log p(x_t)}}_{\text{Standard diffusion model term}} + \underbrace{\boxed{\nabla \log p(y|x_t)}}_{\text{Addition for conditional generation}}$$

Classifier Guidance Generation: How does it work?

- Consider the gradient $\gamma \nabla_{x_t} \log p(y|x_t)$, where γ is a guidance scale on importance for classifier guidance
- Revert the gradient and the logarithm operations to see what's going on:

$$p_\gamma(x_t|y) \propto p(x_t) \cdot p(y|x_t)^\gamma$$

Classifier Guidance Generation: How does it work?

- Consider the gradient $\gamma \nabla_{x_t} \log p(y|x_t)$, where γ is a guidance scale on importance for classifier guidance
- Revert the gradient and the logarithm operations to see what's going on:

$$p_\gamma(x_t|y) \propto p(x_t) \cdot p(y|x_t)^\gamma$$

- Raising the conditional distribution term to power of γ corresponds to tuning the temperature of that distribution, where γ is the inverse temperature parameter

Classifier Guidance Generation: How does it work?

- Consider the gradient $\gamma \nabla_{x_t} \log p(y|x_t)$, where γ is a guidance scale on importance for classifier guidance
- Revert the gradient and the logarithm operations to see what's going on:

$$p_\gamma(x_t|y) \propto p(x_t) \cdot p(y|x_t)^\gamma$$

- Raising the conditional distribution term to power of γ corresponds to tuning the temperature of that distribution, where γ is the inverse temperature parameter
- Classifier guidance thus enables us to apply temperature tuning selectively to the segment of the distribution that reflects the impact of the conditioning signal

Classifier Guidance Algorithm

Algorithm 2 Classifier guided DDIM sampling, given a diffusion model $\epsilon_\theta(x_t)$, classifier $p_\phi(y|x_t)$, and gradient scale s.

Input: class label y, gradient scale s
$x_T \leftarrow$ sample from $\mathcal{N}(0, \mathbf{I})$
for all t from T to 1 **do**
  $\hat{\epsilon} \leftarrow \epsilon_\theta(x_t) - \sqrt{1-\bar{\alpha}_t}\,\nabla_{x_t} \log p_\phi(y|x_t)$
  $x_{t-1} \leftarrow \sqrt{\bar{\alpha}_{t-1}}\left(\frac{x_t - \sqrt{1-\bar{\alpha}_t}\hat{\epsilon}}{\sqrt{\bar{\alpha}_t}}\right) + \sqrt{1-\bar{\alpha}_{t-1}}\hat{\epsilon}$
end for
return x_0

Classifier Guidance Generation: Sample Outputs

$\gamma = 1.0$

$\gamma = 10.0$

Classifier guidance to condition on class "*Pembroke Welsh corgi*". Using $\gamma = 1.0$ does not produce convincing samples, whereas a guidance scale of 10.0 produces more class-consistent images

Credit: Dhariwal et al, Diffusion Models Beat GANs on Image Synthesis, NeurIPS 2021

Classifier Guidance Generation: Limitations

- Requires a dedicated classifier specifically for guidance
- Classifier used for guidance also needs to tackle high noise levels to provide a useful signal
- Classifier gradients are imperfect and yield arbitrary directions in input space

Classifier-Free Guidance Generation

How to get class guidance without an independent classifier?

Ho et al, Classifier-Free Diffusion Guidance, NeurIPS Workshop 2021

Classifier-Free Guidance Generation

How to get class guidance without an independent classifier?

- Use the diffusion model itself to get perturbations for guidance!

Ho et al, Classifier-Free Diffusion Guidance, NeurIPS Workshop 2021

Classifier-Free Guidance Generation

How to get class guidance without an independent classifier?

- Use the diffusion model itself to get perturbations for guidance!
- From Bayes rule:

$$p(y|x_t) = \frac{p(x_t \mid y) \cdot p(y)}{p(x_t)}$$

$$\log p(y \mid x_t) = \log p(x_t \mid y) + \log p(y) - \log p(x_t)$$

$$\nabla_{x_t} \log p(y|x_t) = \nabla_{x_t} \log p(x_t|y) - \nabla_{x_t} \log p(x_t)$$

Ho et al, Classifier-Free Diffusion Guidance, NeurIPS Workshop 2021

Classifier-Free Guidance Generation

How to get class guidance without an independent classifier?

- Use the diffusion model itself to get perturbations for guidance!
- From Bayes rule:

$$p(y|x_t) = \frac{p(x_t \mid y) \cdot p(y)}{p(x_t)}$$

$$\log p(y \mid x_t) = \log p(x_t \mid y) + \log p(y) - \log p(x_t)$$

$$\nabla_{x_t} \log p(y|x_t) = \nabla_{x_t} \log p(x_t|y) - \nabla_{x_t} \log p(x_t)$$

- A few slides ago, we had:

$$\nabla_{x_t} \log p(x_t|y) = \nabla \log \left(\frac{p(x_t)p(y|x_t)}{p(y)} \right) = \nabla \log p(x_t) + \gamma \nabla \log p(y|x_t)$$

Ho et al, Classifier-Free Diffusion Guidance, NeurIPS Workshop 2021

Classifier-Free Guidance Generation

Susbtituting, we get:

$$
\begin{aligned}
\nabla_{x_t} \log p(x_t|y) &= \nabla \log p(x_t) + \gamma \Big(\nabla_{x_t} \log p(x_t|y) - \nabla_{x_t} \log p(x_t) \Big) && (1) \\
&= \gamma \nabla_{x_t} \log p(x_t|y) - (\gamma - 1) \nabla_{x_t} \log p(x_t) && (2) \\
&\approx \gamma \overset{\text{Conditional score}}{\boxed{s_\theta(s_t, y, t)}} - (\gamma - 1) \overset{\text{Unconditional score}}{\boxed{s_\theta(s_t, \emptyset, t)}} && (3) \\
&= \overset{\text{Unconditional score}}{\boxed{s_\theta(s_t, \emptyset, t)}} + \gamma \Big(\overset{\text{Conditional score}}{\boxed{s_\theta(s_t, y, t)}} - \overset{\text{Unconditional score}}{\boxed{s_\theta(s_t, \emptyset, t)}} \Big) && (4)
\end{aligned}
$$

Same neural network can generate both unconditional and conditional scores!

Classifier-Free Guidance Generation

Susbtituting, we get:

$$\nabla_{x_t} \log p(x_t|y) = \nabla \log p(x_t) + \gamma\Big(\nabla_{x_t} \log p(x_t|y) - \nabla_{x_t} \log p(x_t)\Big) \quad (1)$$

$$\approx \gamma \underbrace{\boxed{s_\theta(s_t, y, t)}}_{\text{Conditional score}} - (\gamma - 1) \underbrace{\boxed{s_\theta(s_t, \emptyset, t)}}_{\text{Unconditional score}} \quad (3)$$

$$= \underbrace{\boxed{s_\theta(s_t, \emptyset, t)}}_{\text{Unconditional score}} + \gamma\Big(\underbrace{\boxed{s_\theta(s_t, y, t)}}_{\text{Conditional score}} - \underbrace{\boxed{s_\theta(s_t, \emptyset, t)}}_{\text{Unconditional score}} \Big) \quad (4)$$

Same neural network can generate both unconditional and conditional scores!

Classifier-Free Guidance Generation

Susbtituting, we get:

$$\nabla_{x_t} \log p(x_t|y) = \nabla \log p(x_t) + \gamma \left({\color{red}\nabla_{x_t} \log p(x_t|y)} - {\color{blue}\nabla_{x_t} \log p(x_t)} \right) \tag{1}$$

$$= \overset{\text{Unconditional score}}{\boxed{s_\theta(s_t, \emptyset, t)}} + \gamma \left(\overset{\text{Conditional score}}{\boxed{s_\theta(s_t, y, t)}} - \overset{\text{Unconditional score}}{\boxed{s_\theta(s_t, \emptyset, t)}} \right) \tag{4}$$

Same neural network can generate both unconditional and conditional scores!

Classifier-Free Guidance Generation

Susbtituting, we get:

$$\nabla_{x_t} \log p(x_t|y) = \nabla \log p(x_t) + \gamma\left(\nabla_{x_t}\log p(x_t|y) - \nabla_{x_t}\log p(x_t)\right) \tag{1}$$

(4)

Same neural network can generate both unconditional and conditional scores!

Classifier-Free Guidance Generation

Susbtituting, we get:

$$\nabla_{x_t} \log p(x_t|y) = \nabla \log p(x_t) + \gamma\Big(\nabla_{x_t} \log p(x_t|y) - \nabla_{x_t} \log p(x_t)\Big) \tag{1}$$

$$= \gamma \nabla_{x_t} \log p(x_t|y) - (\gamma - 1)\nabla_{x_t} \log p(x_t) \tag{2}$$

$$\approx \gamma \overset{\text{Conditional score}}{\boxed{s_\theta(s_t, y, t)}} - (\gamma - 1) \overset{\text{Unconditional score}}{\boxed{s_\theta(s_t, \emptyset, t)}} \tag{3}$$

$$= \overset{\text{Unconditional score}}{\boxed{s_\theta(s_t, \emptyset, t)}} + \gamma\Big(\overset{\text{Conditional score}}{\boxed{s_\theta(s_t, y, t)}} - \overset{\text{Unconditional score}}{\boxed{s_\theta(s_t, \emptyset, t)}} \Big) \tag{4}$$

Same neural network can generate both unconditional and conditional scores!

Ho et al, Classifier-Free Diffusion Guidance, NeurIPS Workshop 2021

Classifier-Free Guidance Generation

How to get class guidance without an independent classifier?

Paper: Ho et al, Classifier-Free Diffusion Guidance, NeurIPS Workshop 2021

Classifier-Free Guidance Generation

How to get class guidance without an independent classifier?

- Use the diffusion model itself to get perturbations for guidance!
- Train a conditioned diffusion model $\in_\theta(x_t, t, y)$ with conditioning dropout $\in_\theta(x_t, t, \phi)$, where conditioning information y is removed to make the model unconditional
- In practice, it is often replaced with a special input value representing the absence of conditioning information
- Both conditioned and unconditioned diffusion models can be learned via a single neural network

Paper: Ho et al, Classifier-Free Diffusion Guidance, NeurIPS Workshop 2021

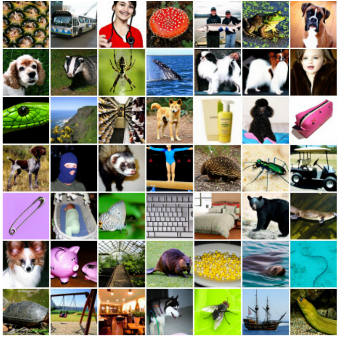
Classifier-Free Guidance Algorithm

Algorithm 1 Joint training a diffusion model with classifier-free guidance

Require: p_{uncond}: probability of unconditional training

1: **repeat**
2: $(\mathbf{x}, \mathbf{c}) \sim p(\mathbf{x}, \mathbf{c})$ ▷ Sample data with conditioning from the dataset
3: $\mathbf{c} \leftarrow \varnothing$ with probability p_{uncond} ▷ Randomly discard conditioning to train unconditionally
4: $\lambda \sim p(\lambda)$ ▷ Sample log SNR value
5: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
6: $\mathbf{z}_\lambda = \alpha_\lambda \mathbf{x} + \sigma_\lambda \boldsymbol{\epsilon}$ ▷ Corrupt data to the sampled log SNR value
7: Take gradient step on $\nabla_\theta \left\| \boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda, \mathbf{c}) - \boldsymbol{\epsilon} \right\|^2$ ▷ Optimization of denoising model
8: **until** converged

Classifier-Free Guidance Generation: Sample Outputs



$\gamma = 0.0$

$\gamma = 1.0$

For $\gamma = 0$, we recover the unconditional model, and for $\gamma = 1$, we get the standard conditional model

Classifier-Free Guidance vs Classifier Guidance

Classifier-free guidance seems to work better than classifier guidance. Why?

Classifier-Free Guidance vs Classifier Guidance

Classifier-free guidance seems to work better than classifier guidance. Why?

- Key reason is that we have constructed the "classifier" from the generative model itself!
- Generative model gradients are much more robust

Classifier-Free Guidance vs Classifier Guidance

Classifier-free guidance seems to work better than classifier guidance. Why?

- Key reason is that we have constructed the "classifier" from the generative model itself!
- Generative model gradients are much more robust
- We have to train a single generative model, and conditioning dropout is trivial to implement

Homework

Readings

- Lilian Weng, What are Diffusion Models?
- (YouTube video) Tutorial on Denoising Diffusion-based Generative Modeling: Foundations and Applications
- Papers referenced in the initial slide