TECHNION
Israel Institute of Technology

DiDiCM: Discrete Diffusion Classification Modeling

Advancing Image Classification with Discrete Diffusion Classification Modeling
CVPR (Oral)


Omer Belhasin


Shelly Golan


Ran El-Yaniv


Michael Elad

Image Classification under Uncertainty

- **Standard classifiers assume a perfect world:** each image maps to a single clean target.
- **In reality, data may be corrupted:** blur, occlusion, and ambiguity might distort this assumption; The true target distribution can shift.
- **With limited data at hand:** This effect becomes even more pronounced.

Corruption

$\Rightarrow$

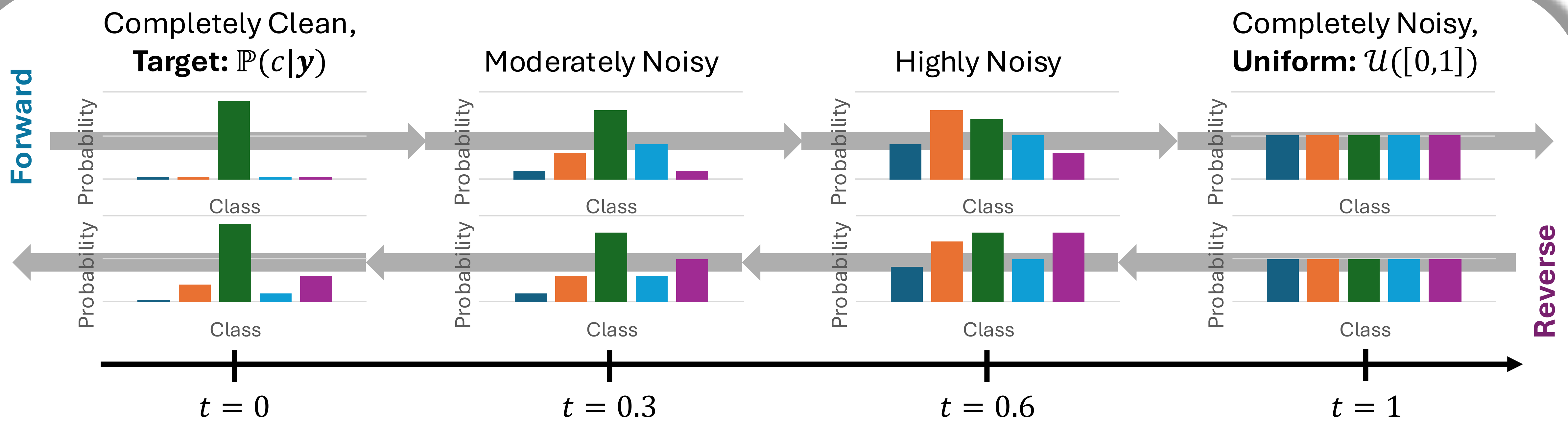
$\boldsymbol{x}, c \sim \mathbb{P}(\boldsymbol{x}, c)$ $\quad$ $\boldsymbol{y} \sim \mathbb{P}(\boldsymbol{y}|\boldsymbol{x})$

Training Target $\quad$ **Effective Target**

$\text{one-hot}\,(c = \text{Cat}) \neq \mathbb{P}(c = \text{Cat} \mid \boldsymbol{y})$

Key Question: Can classification model uncertainty as a generative process over labels?

Core Idea: Diffuse Class Probabilities



- **Core Idea:** Apply continuous-time discrete diffusion process in the class probability space.
- **Benefit**: The model can express multiple plausible classes during denoising rather than committing immediately to one label.

Continuous-Time Markov Chain over Class Prob.

Forward Process

$$\frac{\delta q(c_t|\boldsymbol{y})}{\delta t} = R_t \cdot q(c_t|\boldsymbol{y})$$

- **Mechanism:** Linear ODE with uniform transition rate matrix.
- **Solution:** Closed-form solution via analytical derivation.
- **Outcome:** Increases entropy until distribution is uniform.

Reverse Process

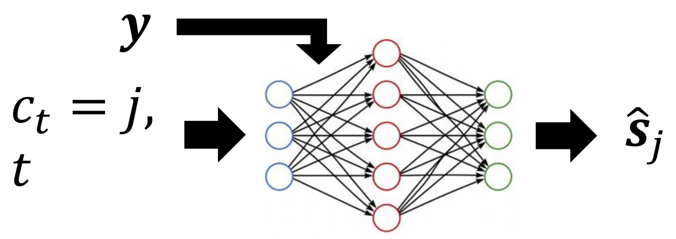
$$\frac{\delta p(c_t|\boldsymbol{y})}{\delta t} = \bar{R}_t^{\theta} \cdot p(c_t|\boldsymbol{y})$$

- **Mechanism:** Linear ODE with learned reversal transition rate matrix.
- **Solution:** Approximated via Euler simulation.
- **Outcome:** Restores signal from noise conditioned on the input image.

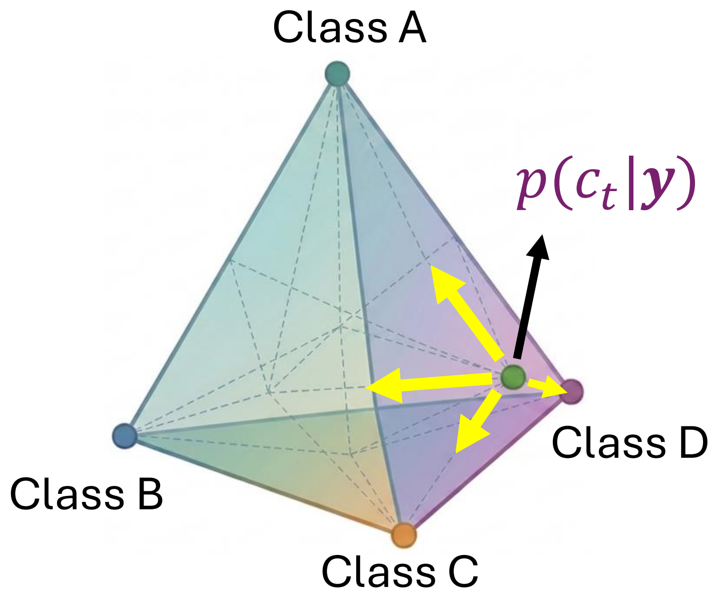
Classification becomes reverse-time denoising in a K-class simplex.

The Model Learns the Concrete Score

- **Our model takes a noisy label as input alongside the image:** it learns how the label distribution transitions across classes.



- **This is driven by the Concrete Score:** a generalization of the continuous score function from Score Matching.



$\Rightarrow$ Predicting the **Concrete Score**

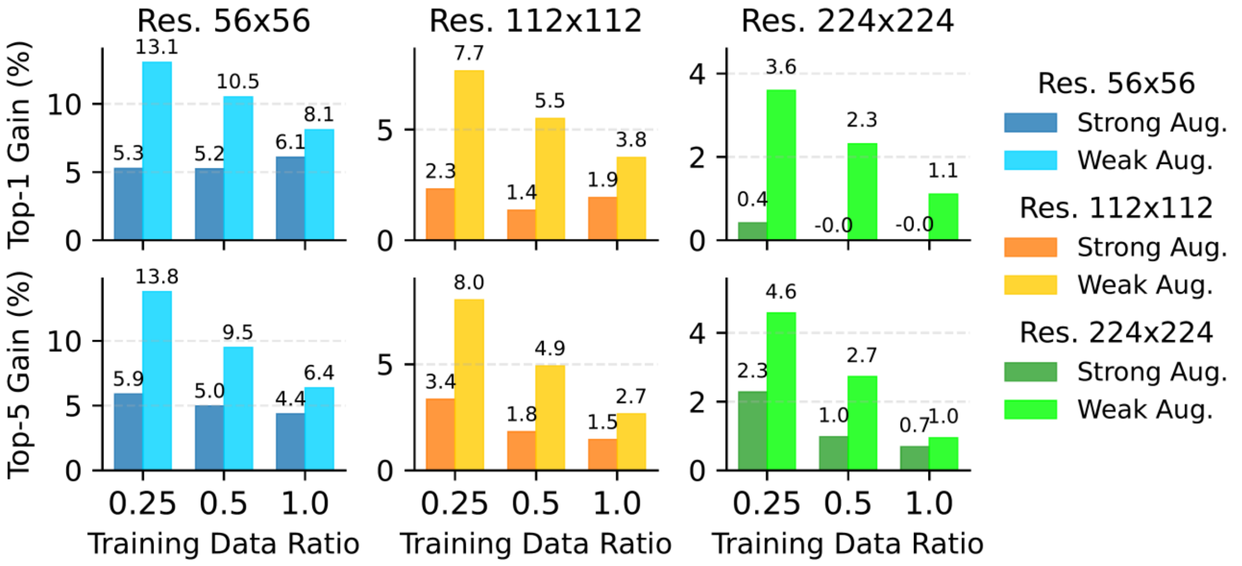
$$[\hat{\boldsymbol{s}}_j]_i \approx \frac{q(c_t = i|\boldsymbol{y})}{q(c_t = j|\boldsymbol{y})}$$

Meng, et al. "Concrete score matching: Generalized score matching for discrete data". (2022)

Lou, et al. "Discrete diffusion modeling by estimating the ratios of the data distribution". (2023)

Stronger Behavior as Uncertainty Increases

- **In high uncertainty:** DiDiCM surpasses conventional classifiers, exceeding SOTA top-5 accuracy.
- **In low uncertainty:** DiDiCM matches SOTA top-1 accuracy while still surpassing SOTA top-5 accuracy.
- **A clear trend emerges: the harder the setting, the larger the gap;** Illustrating robustness where classifiers struggle most.



Main comparison: **DiDiCM** vs. **ResNet50** using state-of-the-art training policy.

TECHNION
Israel Institute of Technology

Thank you

Advancing Image Classification with Discrete Diffusion Classification Modeling

Corrupted Image
$\boldsymbol{y}$

DiDiCM $\longrightarrow$ Generative Classification $p_\theta(c|\boldsymbol{y})$

Talk with me: omerb01@gmail.com

Codebase: https://github.com/omerb01/didicm