Class-Conditional Vial Image Generation Using Diffusion Models for Enhanced Visual Inspection

Zhengqi Gao[1], Stefania Vitillo[2], Kevin Joinson[2], Duane S. Boning[1]

[1]Massachusetts Institute of Technology, [2]Takeda Inc

Contact: zhengqi@mit.edu, stefania.vitillo@takeda.com

Abstract

- Approximately 50% of the global population lives in regions affected by vector-borne diseases, with hundreds of millions of cases reported annually. These diseases can be severe, and the risk of complications increases with repeated infections. Vaccination offers a promising preventive measure, particularly in the absence of targeted treatments.
- Lyophilization is a crucial step in making vaccine accessible due to its thermal instability in liquid form. However, automated inspection of lyophilized product which is required to meet global demand, has a high false reject rate of up to 30%.
- Automated visual inspection (AVI) seeks to develop and evaluate novel approaches (signal processing, AI-enabled) for inspecting defective products in pharmaceutical pharmaceutical manufacturing.
- Previous works show with transfer learning (fine-tuning a VGG16 model with a few vial images), it is possible to reduce false positive rate to less than 5%.
- A key challenge in AVI is obtaining labeled data. While raw images are accessible, labeling them is time-consuming. However, we need thousands labeled data to train a neural network. Defective samples are rare, and when a production line isn't in routine operation, acquiring new images is harder.
- Ultimate long-term goal: **Lower the re-inspection rate by introducing an AI-enabled vaccine inspection approach**

Background

Dual-Threshold Classification

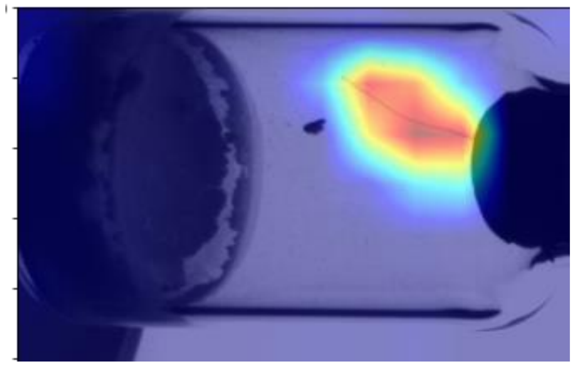


Fig. 2:GradCam Visualization

- Pretrained VGG16 finetuned achieve 92% accuracy on a proprietary Takeda vial dataset
- Dual-threshold reduces FP rate; send uncertain to expert
- GradCAM used to visualize which region the model uses to make classification decision

Label Error Learning

In real world pharmaceutical problems, potentially multiple noisy labels are available for one vial image. We propose a new label error learning algorithm to improve classification accuracy, with the key learning sample-dependent weight vector and confusion matrices [1].

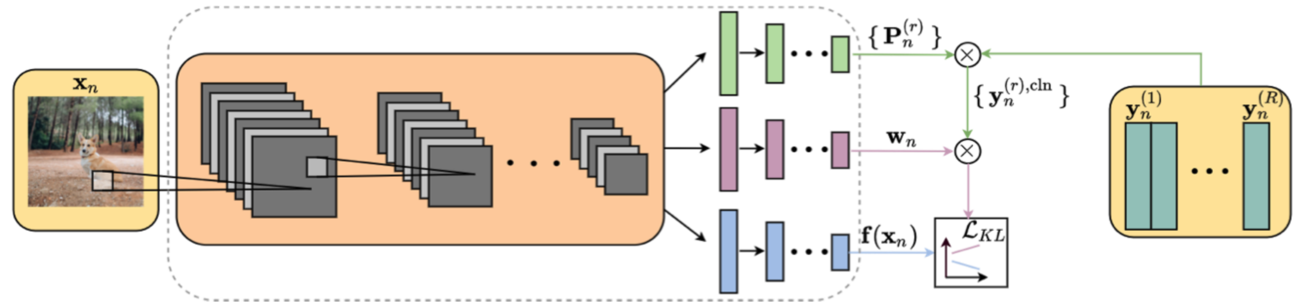


Fig. 3: Training flow of our proposed method

	$\epsilon = 580$	$\epsilon = 600$	$\epsilon = 620$	$\epsilon = 650$	$\epsilon = 680$
w/ AnT-1's	38.36	34.25	28.24	22.71	15.81
w/ AnT-2's	34.72	30.75	24.50	17.91	12.70
w/ AnT-3's	56.35	52.62	47.16	40.41	34.86
Mjv	53.83	46.67	41.13	35.90	29.83
TraceReg	58.51	55.13	48.90	43.07	36.54
MBEM	45.18	39.12	32.90	26.14	20.06
WDN	56.03	51.86	46.90	37.94	31.65
Ours	**73.34**	**70.59**	**68.91**	**62.72**	**54.93**
w/ true label	74.26	74.26	74.26	74.26	74.26

Tab. 1: Compare our method with several other baselines at different annotator error ϵ

Diffusion Models

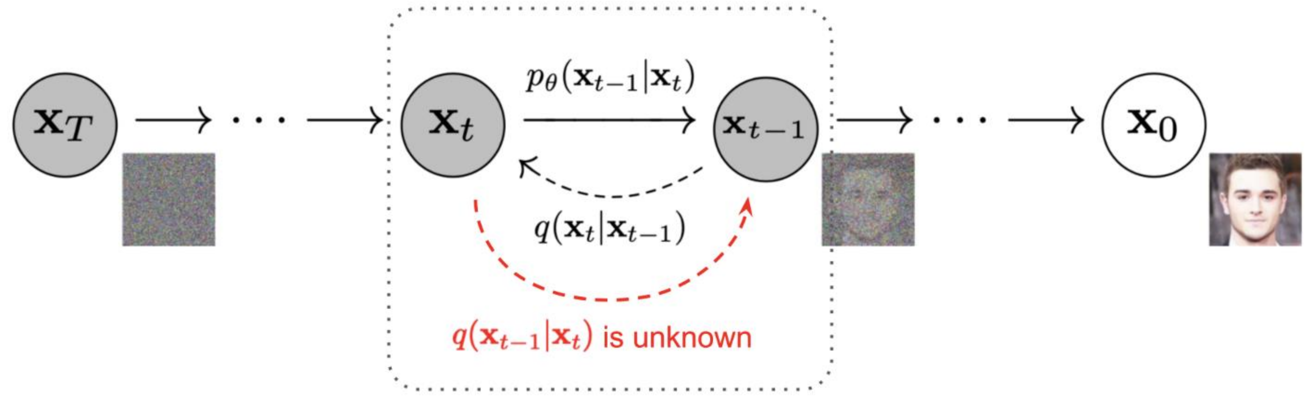


Fig. 4: Denoising diffusion probabilistic model (DDPM). [2]

Forward noising

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

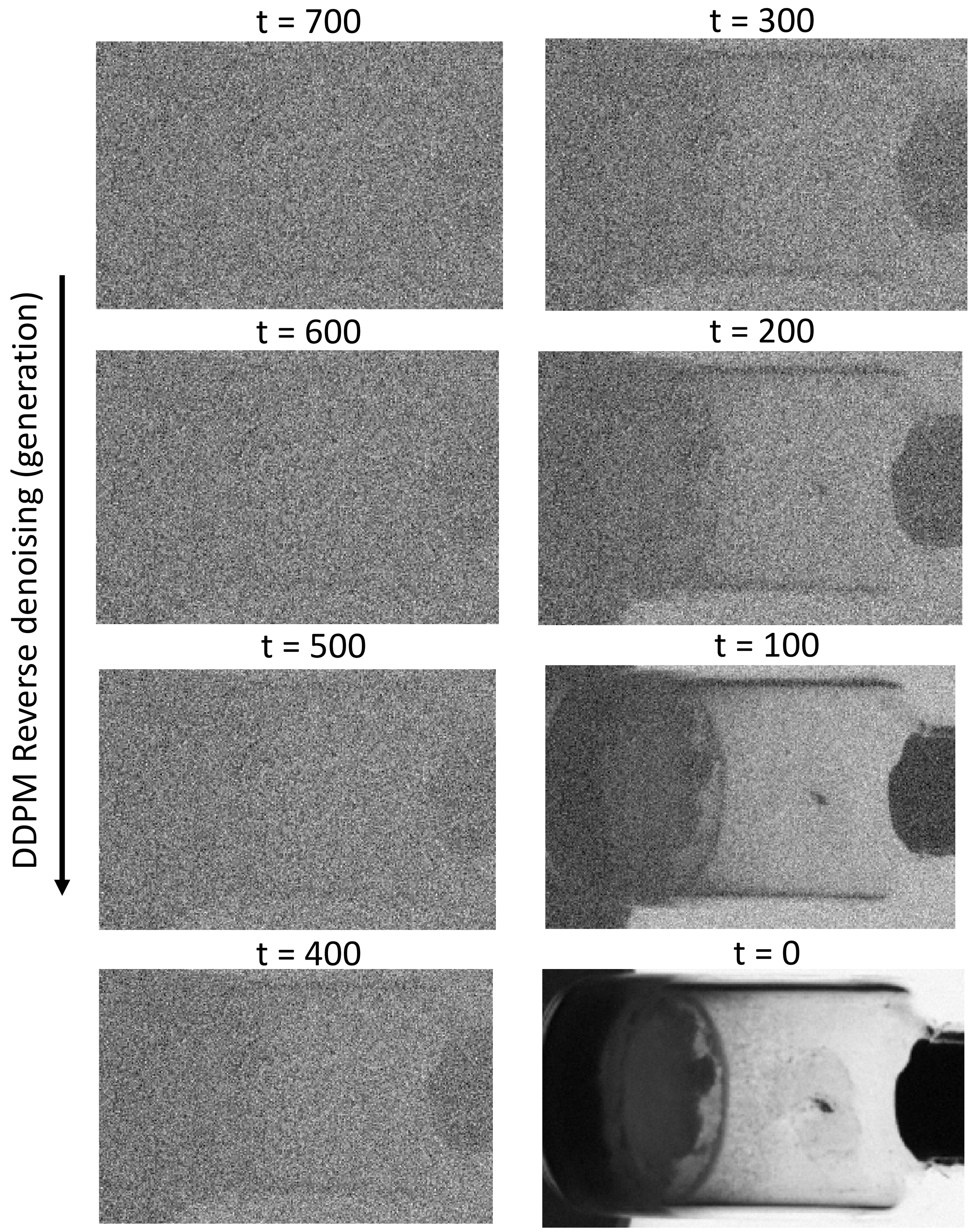
$$q(\mathbf{x}_{0:T}|\mathbf{y}) = q(\mathbf{x}_0|\mathbf{y})\prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}),$$

Reverse denoising

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{y}) = \mathcal{N}(\mathbf{x}_{t-1}|\boldsymbol{\mu}_{\theta,t}, \sigma_t^2\mathbf{I}),$$

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

- Reverse denoising starts from a random Gaussian noise $\mathbf{x}_T \sim N(\mathbf{0}, \mathbf{I})$, gradually denoise to an image $\mathbf{x}_0$.
- In reverse denoising, the transition from $\mathbf{x}_t$ to $\mathbf{x}_{t-1}$ is determined by $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, y)$, which is parametrized by a neural network.
- Forward noising is a constructed target, used to learn the reverse denoising process (parametrized by θ)



Conclusion

AIs help in lyophilized product inspection and can greatly save the development and operation cost. Prior AI models have been developed for inspecting defective vial images. In this work, we further reduce the data acquisition cost, by showing a diffusion model can be trained as an auxiliary source for data. In the future, we attempt to compressively evaluate the boost of using diffusion models in help of AVI.

References

[1] Zhengqi Gao, et al., *Learning from Multiple Annotator Noisy Labels via Sample-Wise Label Fusion*. In European Conference on Computer Vision (ECCV) Oct. 2022.

[2]Jonathan Ho, Ajay Jain, Pieter Abbeel. *Denoising Diffusion Probabilistic Models*. Neurips 2020.