M3D: Dataset Condensation by Minimizing Maximum Mean Discrepancy

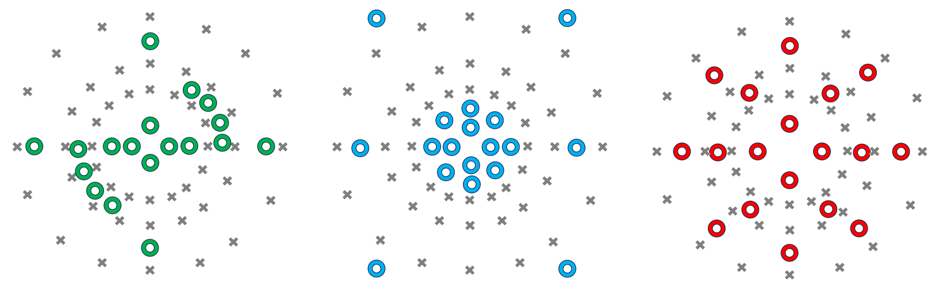
Hansong Zhang, Shikun Li, Pengju Wang, Dan Zeng, Shiming Ge

Summary

Background:
Dataset condensation (also known as dataset distillation) is a process aimed at addressing the challenges associated with the storage and time requirements of training state-of-the-art deep models. It involves creating a small synthetic dataset that retains the essential information of the original large-scale dataset.

Motivation:

- Traditional bi-level optimization used in dataset distillation, while effective, are often impractical for larger datasets due to their computational complexity and inefficiency.
- Distribution-Matching (DM) methods focus on aligning the feature distributions of synthetic and real data, offering greater efficiency but producing less informative examples compared to Optimization-Oriented methods.
- Previous DM-based methods only aligns the first-order moment of the synthetic and real data, which is not sufficient for matching two distributions.



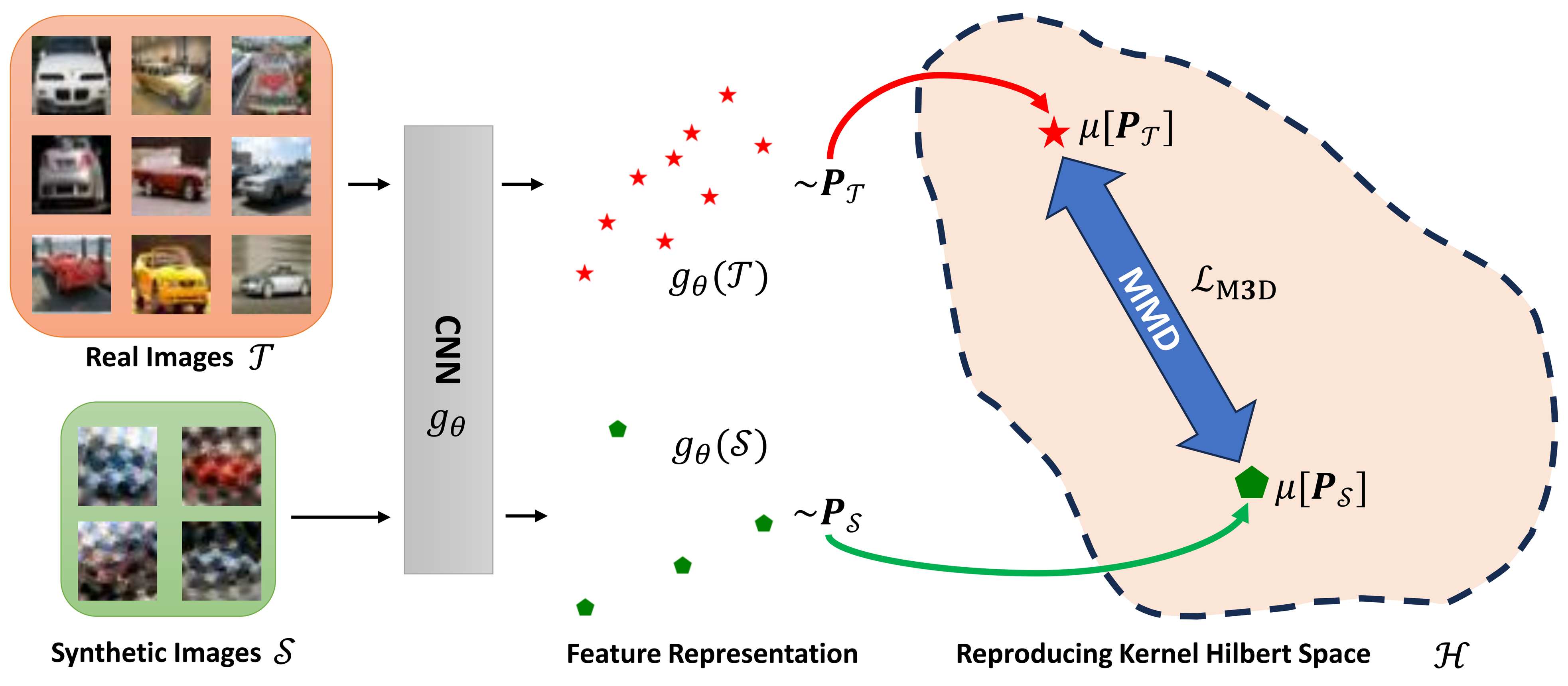
2nd-order Misalignment   3rd-order Misalignment   Aligned Distributions

Contribution:

- We introduce a novel DM-based method named M3D for dataset condensation. This method focuses on minimizing the Maximum Mean Discrepancy between feature representations of synthetic and real images, which is achieved by embedding their distributions in a reproducing kernel Hilbert space.
- Unlike previous DM methods that primarily align only the first moment of distributions, M3D aligns all orders of moments of the distributions of real and synthetic images.
- Extensive experiments are conducted on both low-resolution and high-resolution datasets, where the results indicate the superiority of M3D.

Methodology

Main idea: The core idea of M3D is (1) extract the feature representations of both synthetic and real data; (2) based on classical kernel method, further embed the feature representations to a reproducing kernel Hilbert space, where we can map the infinite order of moments into a form of kernel function.



Reproducing Kernel Hilbert Space (RKHS): Given a kernel $\mathcal{K}$, $\mathcal{H}$ is a Hilbert space of functions $\mathcal{X} \to \mathbb{R}$ with dot product $\langle\cdot,\cdot\rangle$, if $\forall\phi$, satisfying the following properties:

$$\text{Reproducing}: \langle \phi(\cdot), \mathcal{K}(x,\cdot)\rangle = \phi(x).$$

$$\text{Symmetry}: \mathcal{K}(x,x') = \mathcal{K}(x',x)$$

$$\text{Positive}: \mathcal{K}(\cdot,\cdot) \geq 0$$

Previous Distribution-Mathcing: Previous DM methods only align the first-order moment of synthetic and real data, which can be formulated as:

$$\mathcal{S}^\star = \arg\min_{\mathcal{S}} E_{\theta\sim P_\theta} \|\frac{1}{|\mathcal{T}|}\sum_{i=1}^{|\mathcal{T}|} g_\theta(\boldsymbol{x}_i) - \frac{1}{|\mathcal{S}|}\sum_{j=1}^{|\mathcal{S}|} g_\theta(\boldsymbol{s}_j)\|^2$$

Embedding feature repersentations into RKHS: Our M3D further embeds the feature representations into a Reproducing Kernel Hilbert Space, where we can effectively align all order of moments between synthetic and real data:

$$\mathcal{L}_{\text{M3D}} = \hat{\text{MMD}}^2(\boldsymbol{P}_\mathcal{T}, \boldsymbol{P}_\mathcal{S}) = \hat{\mathcal{K}}_{\mathcal{T},\mathcal{T}} + \hat{\mathcal{K}}_{\mathcal{S},\mathcal{S}} - 2\hat{\mathcal{K}}_{\mathcal{T},\mathcal{S}}$$

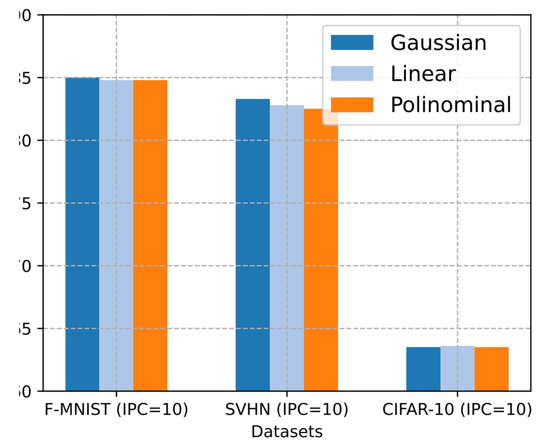
where $\hat{\mathcal{K}}_{X,Y} = \frac{1}{|X|\cdot|Y|}\sum_{i=1}^{|X|}\sum_{j=1}^{|Y|}\mathcal{K}(g_\theta(x_i), g_\theta(y_j))$ with $\{x_i\}_{i=1}^{|X|} \sim X, \{y_j\}_{j=1}^{|Y|} \sim Y.$

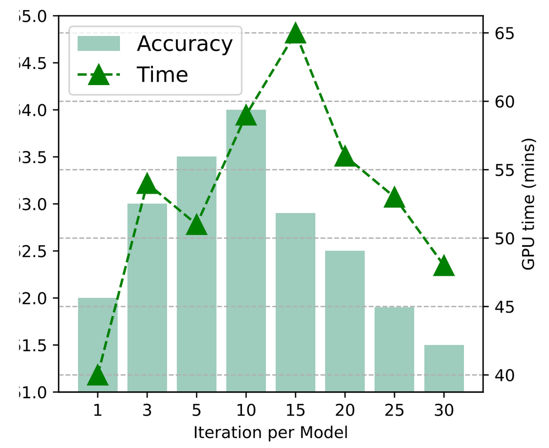
Experiments

Experiments on low-resolution datasets:

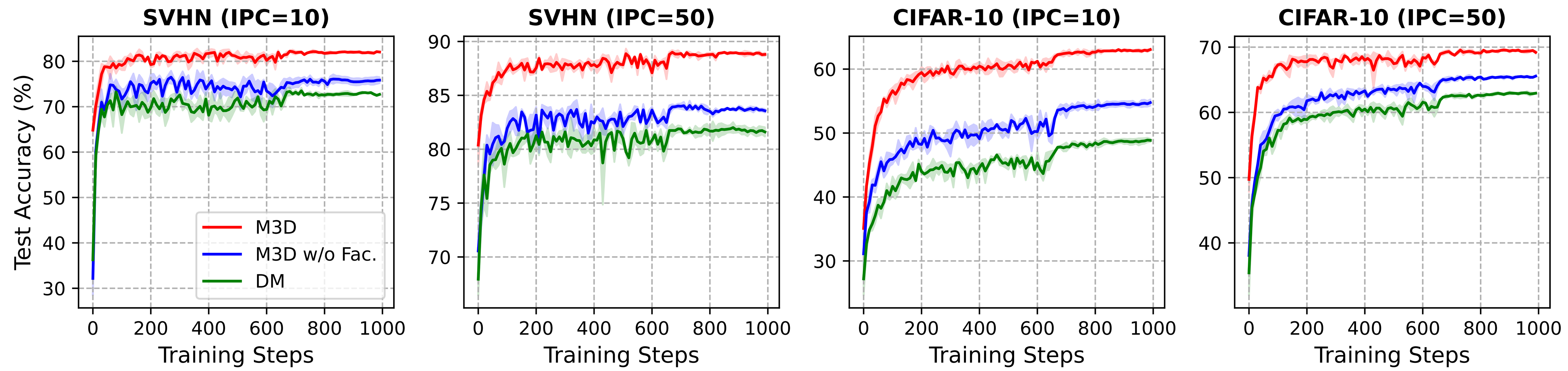
Dataset	IPC	Ratio (%)	Coreset Selection: Random	Herding	K-Center	Dataset Condensation: DC	DSA	CAFE	CAFE+DSA	DM	IDM	M3D	Whole
MNIST	1	0.017	$64.9_{\pm3.5}$	$89.2_{\pm1.6}$	$89.3_{\pm1.5}$	$91.7_{\pm0.5}$	$88.7_{\pm0.6}$	$93.1_{\pm0.3}$	$90.8_{\pm0.5}$	$89.7_{\pm0.6}$	-	**94.4**$_{\pm0.2}$	$99.6_{\pm0.0}$
	10	0.17	$95.1_{\pm0.9}$	$93.7_{\pm0.3}$	$84.4_{\pm1.7}$	$97.4_{\pm0.2}$	**97.8**$_{\pm0.1}$	$97.2_{\pm0.2}$	$97.5_{\pm0.1}$	$97.5_{\pm0.1}$	-	$97.6_{\pm0.1}$	
	50	0.83	$97.9_{\pm0.2}$	$94.8_{\pm0.2}$	$97.4_{\pm0.3}$	$98.8_{\pm0.2}$	**99.2**$_{\pm0.1}$	$98.6_{\pm0.2}$	$98.9_{\pm0.2}$	$98.6_{\pm0.1}$	-	$98.2_{\pm0.2}$	
F-MNIST	1	0.017	$51.4_{\pm3.8}$	$67.0_{\pm1.9}$	$66.9_{\pm1.8}$	$70.5_{\pm0.6}$	$70.6_{\pm0.6}$	$77.1_{\pm0.9}$	$73.7_{\pm0.7}$	$70.7_{\pm0.6}^\dagger$	-	**80.7**$_{\pm0.3}$	$93.5_{\pm0.1}$
	10	0.17	$73.8_{\pm0.7}$	$71.1_{\pm0.7}$	$54.7_{\pm1.5}$	$82.3_{\pm0.4}$	$84.6_{\pm0.3}$	$83.0_{\pm0.4}$	$83.0_{\pm0.3}$	$83.5_{\pm0.3}^\dagger$	-	**85.0**$_{\pm0.1}$	
	50	0.83	$82.5_{\pm0.7}$	$71.9_{\pm0.8}$	$68.3_{\pm0.8}$	$83.6_{\pm0.4}$	**88.7**$_{\pm0.2}$	$84.8_{\pm0.4}$	$88.2_{\pm0.3}$	$88.1_{\pm0.6}^\dagger$	-	$86.2_{\pm0.3}$	
SVHN	1	0.014	$14.6_{\pm1.6}$	$20.9_{\pm1.3}$	$21.0_{\pm1.5}$	$31.2_{\pm1.4}$	$27.5_{\pm1.4}$	$42.6_{\pm3.3}$	$42.9_{\pm3.0}$	$30.3_{\pm0.1}^\dagger$	-	**62.8**$_{\pm0.5}$	$95.4_{\pm0.1}$
	10	0.14	$35.1_{\pm4.1}$	$50.5_{\pm3.3}$	$14.0_{\pm1.3}$	$76.1_{\pm0.6}$	$79.2_{\pm0.5}$	$75.9_{\pm0.6}$	$77.9_{\pm0.6}$	$73.5_{\pm0.5}^\dagger$	-	**83.3**$_{\pm0.7}$	
	50	0.7	$70.9_{\pm0.9}$	$72.6_{\pm0.8}$	$20.1_{\pm1.4}$	$82.3_{\pm0.3}$	$84.4_{\pm0.4}$	$81.3_{\pm0.3}$	$82.3_{\pm0.4}$	$82.0_{\pm0.2}^\dagger$	-	**89.0**$_{\pm0.2}$	
CIFAR-10	1	0.02	$14.4_{\pm2.0}$	$21.5_{\pm1.2}$	$21.5_{\pm1.3}$	$28.3_{\pm0.5}$	$28.8_{\pm0.7}$	$30.3_{\pm1.1}$	$31.6_{\pm0.8}$	$26.0_{\pm0.8}$	**45.6**$_{\pm0.7}$	$45.3_{\pm0.3}$	$84.8_{\pm0.1}$
	10	0.2	$26.0_{\pm1.2}$	$31.6_{\pm0.7}$	$14.7_{\pm0.9}$	$44.9_{\pm0.5}$	$52.1_{\pm0.5}$	$46.3_{\pm0.6}$	$50.9_{\pm0.5}$	$48.9_{\pm0.6}$	$58.6_{\pm0.1}$	**63.5**$_{\pm0.2}$	
	50	1	$43.4_{\pm1.0}$	$40.4_{\pm0.6}$	$27.0_{\pm1.4}$	$53.9_{\pm0.5}$	$60.6_{\pm0.5}$	$55.5_{\pm0.6}$	$62.3_{\pm0.4}$	$63.0_{\pm0.4}$	$67.5_{\pm0.1}$	**69.9**$_{\pm0.5}$	
CIFAR-100	1	0.2	$4.2_{\pm0.3}$	$8.4_{\pm0.3}$	$8.3_{\pm0.3}$	$12.8_{\pm0.3}$	$13.9_{\pm0.3}$	$12.9_{\pm0.3}$	$14.0_{\pm0.3}$	$11.4_{\pm0.3}$	$20.1_{\pm0.3}$	**26.2**$_{\pm0.3}$	$56.2_{\pm0.3}$
	10	2	$14.6_{\pm0.5}$	$17.3_{\pm0.3}$	$7.1_{\pm0.2}$	$25.2_{\pm0.3}$	$32.3_{\pm0.3}$	$27.8_{\pm0.3}$	$31.5_{\pm0.2}$	$29.7_{\pm0.3}$	**45.1**$_{\pm0.1}$	$42.4_{\pm0.2}$	
	50	10	$30.0_{\pm0.4}$	$33.7_{\pm0.5}$	$30.5_{\pm0.3}$	-	$42.8_{\pm0.4}$	$37.9_{\pm0.3}$	$42.9_{\pm0.2}$	$43.6_{\pm0.4}$	$50.0_{\pm0.2}$	**50.9**$_{\pm0.7}$	

Ablation studies:

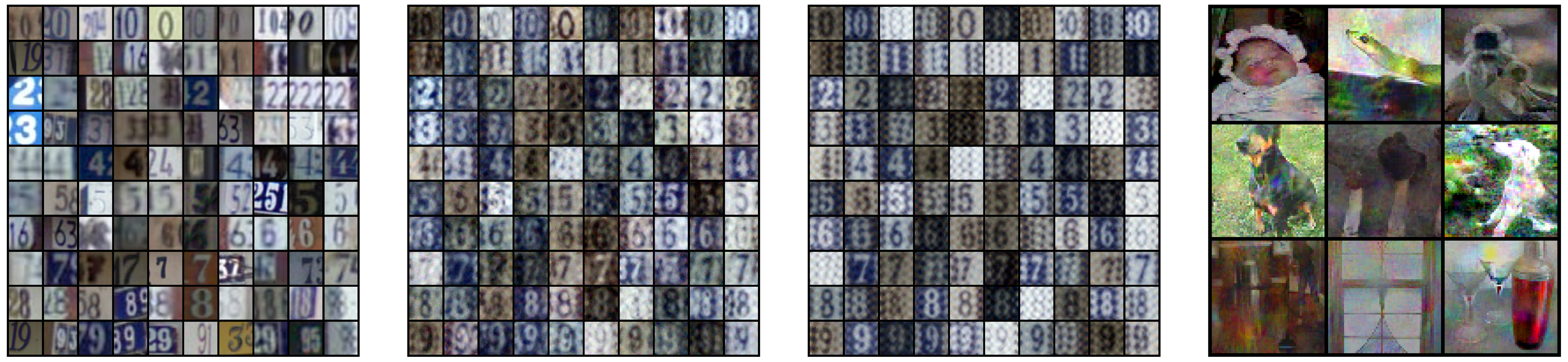




Experiments across varying training steps:



Visualization Results:



Initialized SVHN images   Condensed images by **DM**   Condensed images by **M3D**   Condensed ImageNet-100 images