

# 성장하는 연구원 김현진입니다.

AI RESEARCHER PORTFOLIO



---

Github

<https://github.com/hjinnkim>

Phone

010-4688-1203

Email

khj12203@gmail.com

# 자기소개



## 목표를 향해 성장하는 김현진입니다.

배움에 열린 태도로 다양한 분야를 탐구하며,  
문제 해결에 보람을 느낍니다. 변화에 유연하게  
대응하며 책임감 있게 성장하겠습니다.

### 학력 및 경력 사항

2023.02	UNIST 전기전자공학과 학사 졸업
2025.02	UNIST 인공지능대학원 석사 졸업
Present	UNIST 인공지능대학원 연구원

### 연구 키워드

Generative Models (GANs, Diffusion)  
Generative Model Compression  
Inverse Problem (Super Resolution)

### 논문

1. Singular Value Scaling: Efficient Generative Model Compression via Pruned Weights Refinement (AAAI 2025)
2. BF-STVSR: B-Splines and Fourier-Best Friends for High Fidelity Spatial-Temporal Video Super-Resolution (CVPR 2025)

### 기술 스택



Python



C/C++



Pytorch



Docker

### 언어

한국어 Native

영어 Intermediate (OPIc: IH)

# PROJECTS

1. 생성모델 경량화 연구

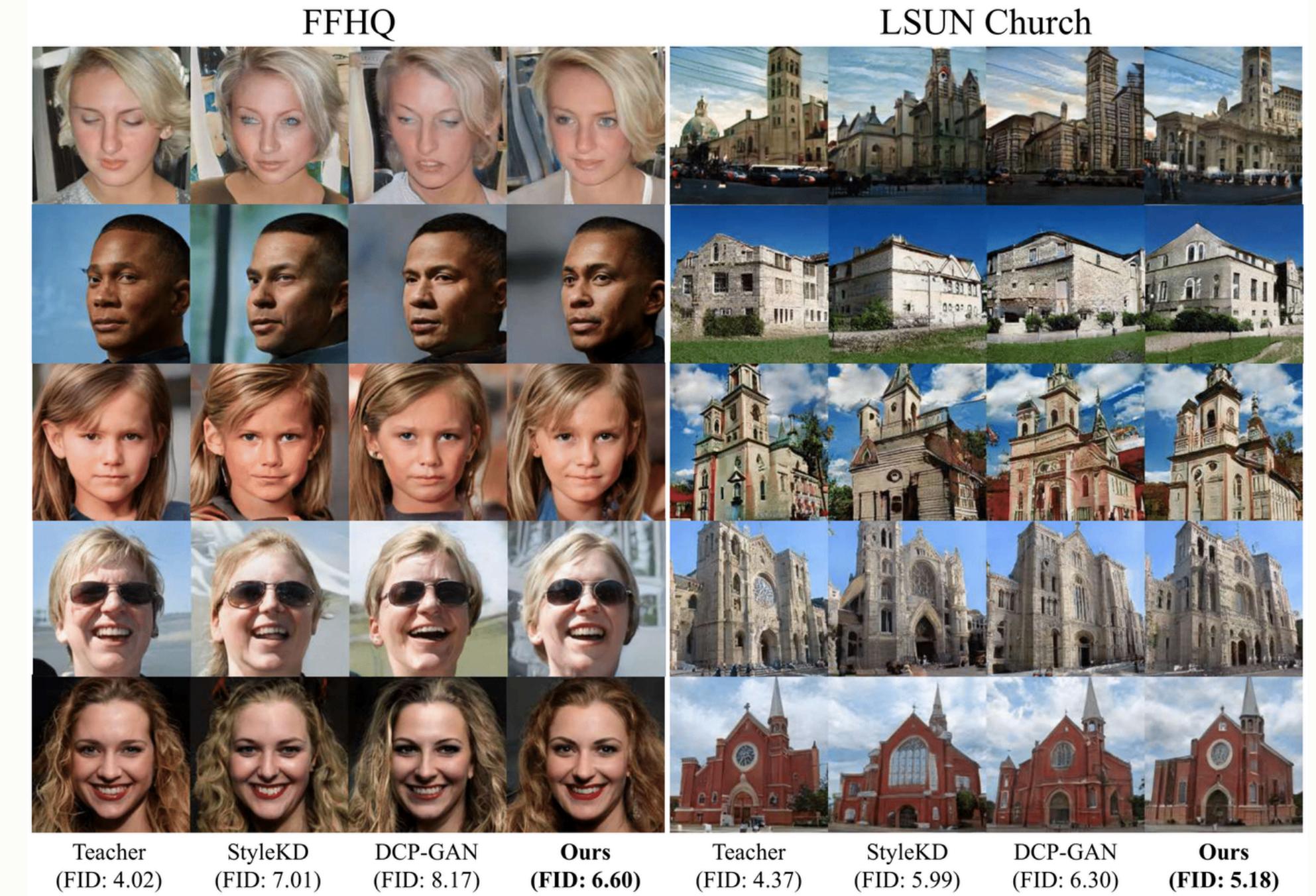
---

2. 시공간 비디오 초해상화 연구

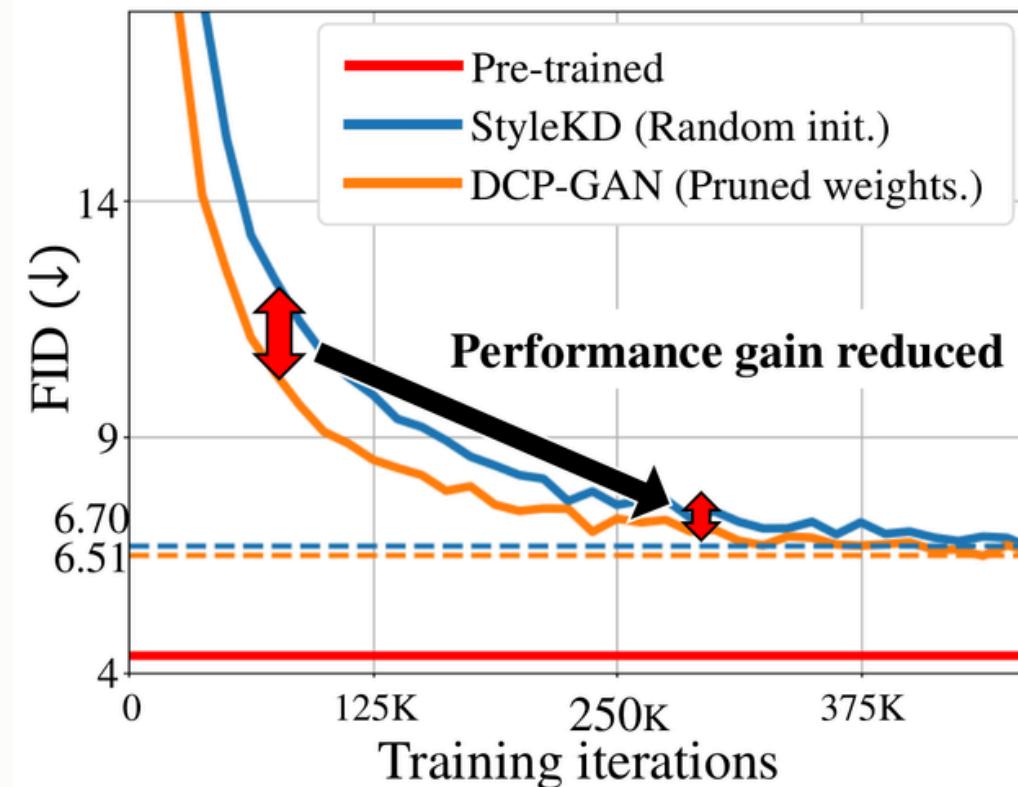
---

# 생성모델 경량화 연구

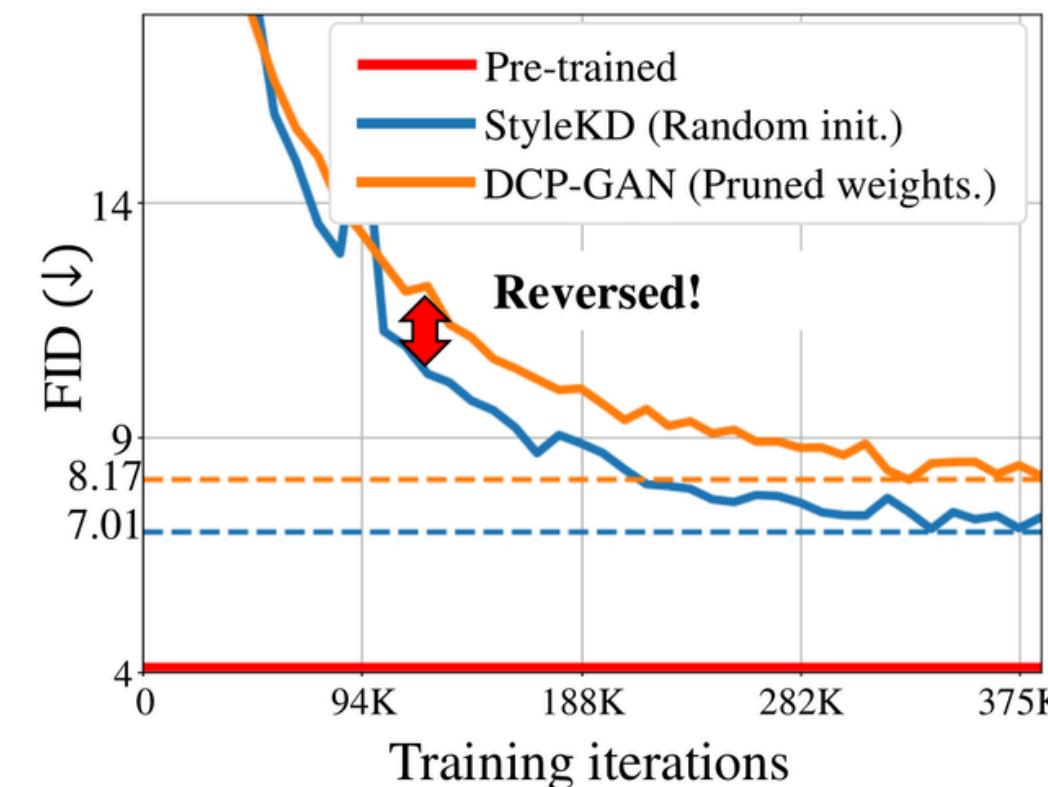
- 생성 모델을 실서비스에 적용하려면 성능을 유지하면서 연산량과 메모리를 줄이는 **모델 압축 (Model Compression)**이 필요합니다.
- 사전 학습 가중치(weight)에서 불필요한 뉴런을 제거하는 **가지치기(pruning)** 기법은 랜덤 초기화 대비 압축된 모델에 좋은 가중치 초기화를 제공하는 것이 목표입니다.
- 기존 연구들은 사전 학습 모델의 중요 연결만 유지하는데 집중해 이어지는 **파인 투닝(fine-tuning)**에 미치는 영향을 간과해 **비효율**을 초래합니다.
- 본 연구에서는 가지치기된 가중치의 **학습 비효율의 원인을 분석**하고, **가중치 정제 (weights refinement)**를 통해 압축된 모델의 **학습 효율 개선**하였습니다.



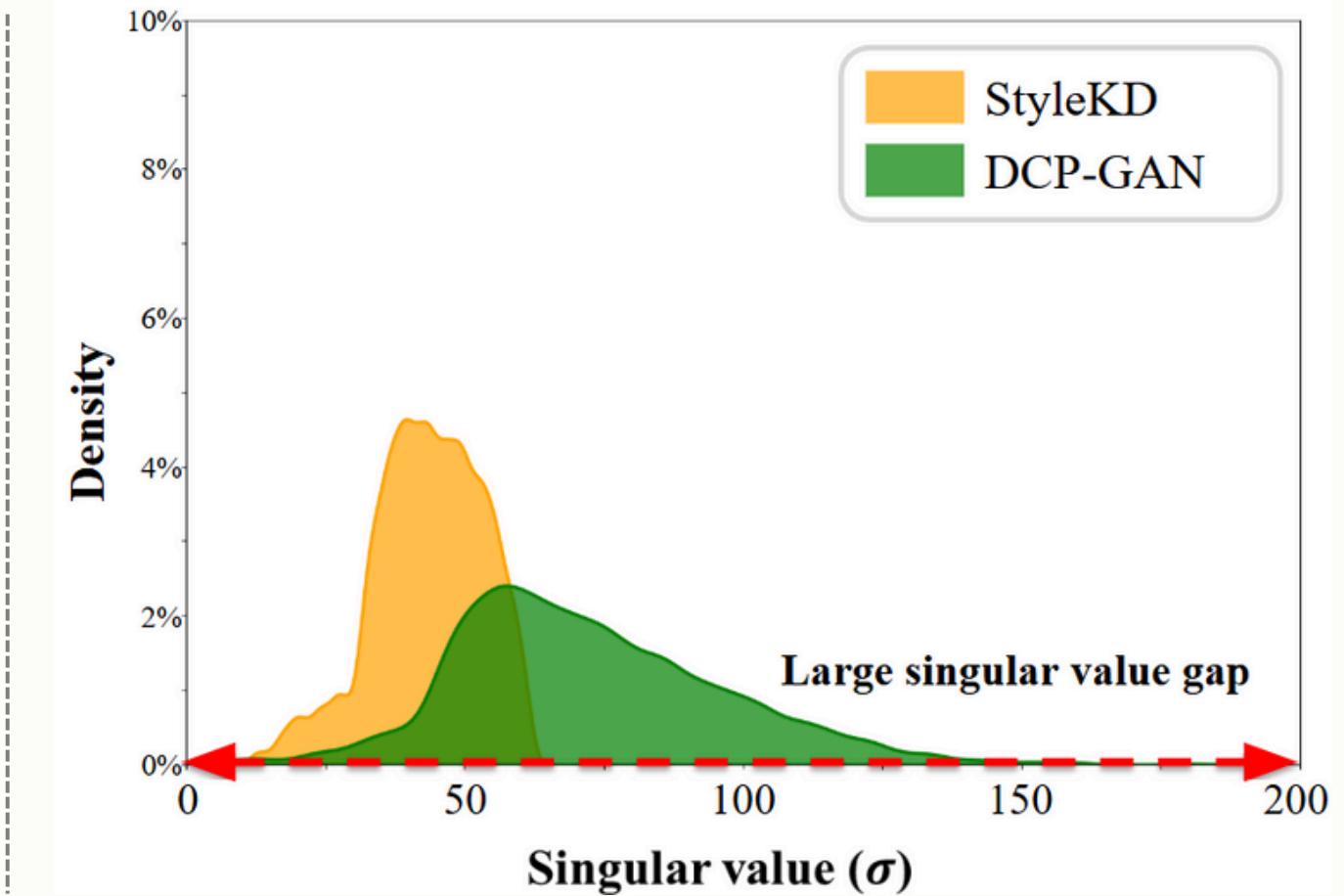
# 생성모델 경량화 연구



(a) FID graph of StyleGAN2 (base)

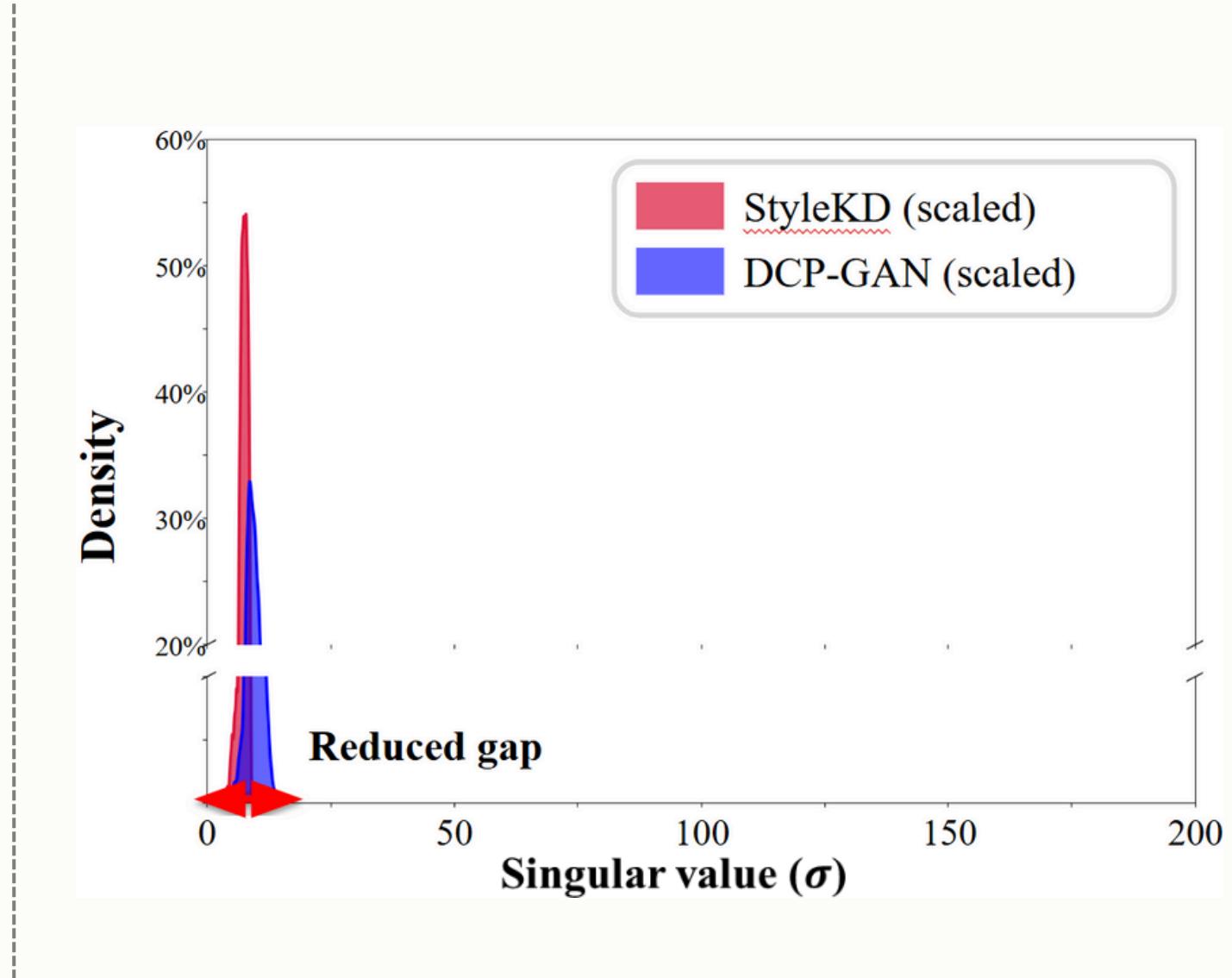
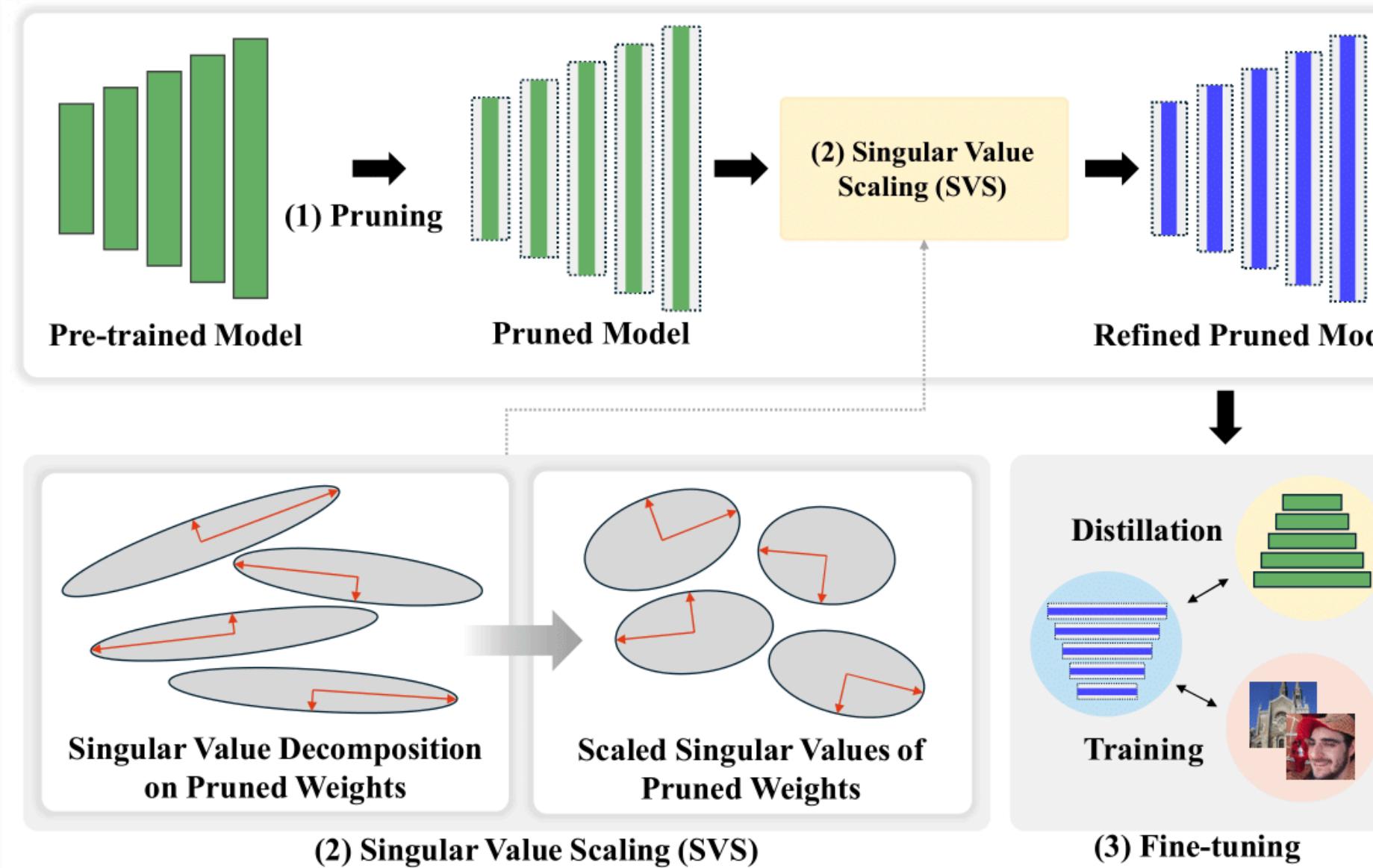


(b) FID graph of StyleGAN2 (small)



- 대표적인 생성모델인 StyleGAN 압축의 경우, 랜덤 초기화와 가지치기된 모델의 **성능 차이**는 파인 튜닝 후반부 **급격히 감소**합니다.
- 모델의 크기가 더 작아질 경우, 가지치기된 모델의 **성능**이 파인 튜닝 초반 랜덤 초기화된 모델에게 **추월** 당합니다.
- 가지치기된 모델의 **학습 효율 저해 요인**을 분석하기 위해 가중치를 특이값 분해(SVD)로 분석하였고, **지배적 특이벡터**의 존재를 확인할 수 있었습니다.
- 가중치 내 지배적 특이벡터는 순/역전파 과정을 해당 특이벡터 방향으로 왜곡시키고, 다른 특이벡터의 영향을 약화(**gradient vanishing**) 시킬 수 있습니다.

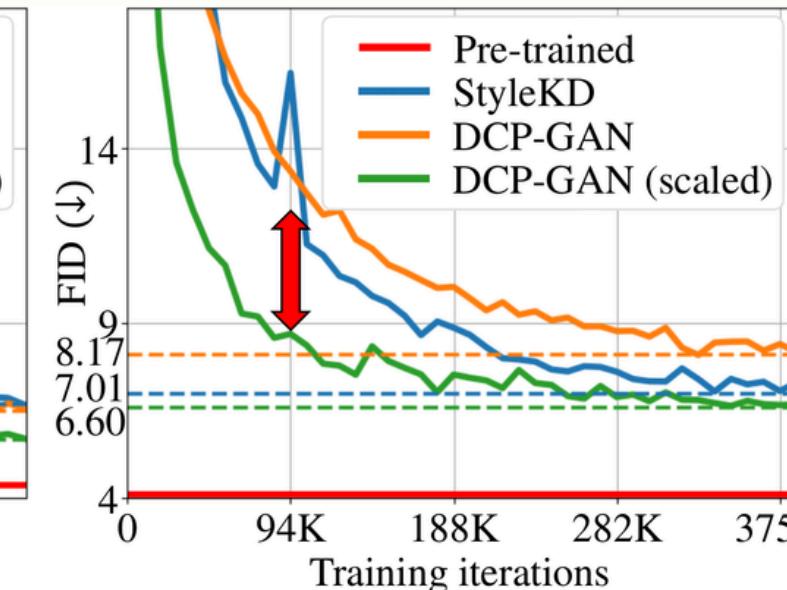
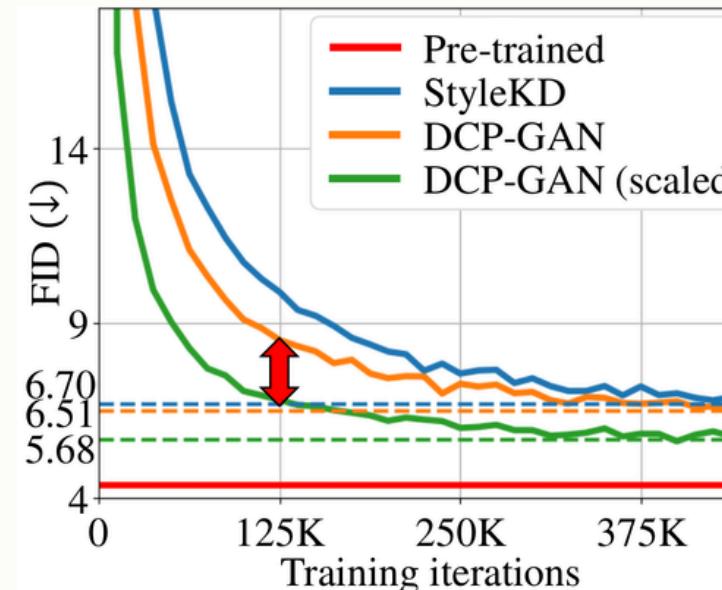
# 생성모델 경량화 연구



- 앞선 분석을 기반으로, 가지치기된 가중치를 초기화 단계에서 정제하는 Singular Value Scaling(SVS)를 제안하였습니다.
- 파인 투닝 이전에 가지치기된 가중치 특이값에 제곱근을 취하여, 각 특이벡터가 학습에 균등하게 기여하도록 만들었습니다.
- 이를 통해 사전 학습된 모델의 지식을 유지하면서, 향상된 파인 투닝 성능을 유지할 수 있었습니다.

# 생성모델 경량화 연구

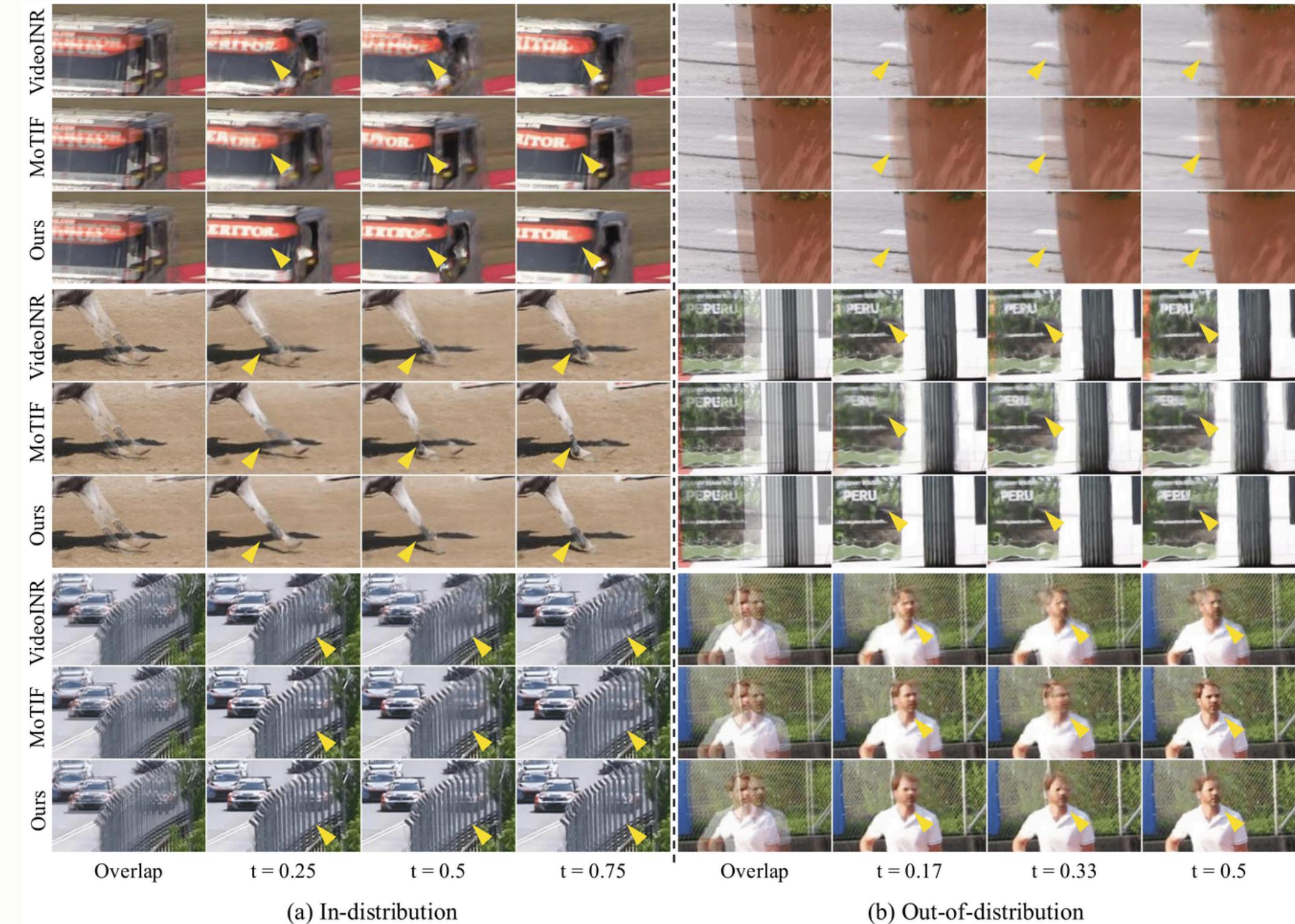
Dataset	Method	Arch	Params ↓	FLOPs ↓	FID ↓	P ↑	R ↑	D ↑	C ↑
FFHQ	Teacher	base	30.0M	45.1B	4.29	0.762	0.561	1.061	0.845
	StyleKD [Xu et al. 2022]				6.70	0.711	0.551	0.879	0.786
	DCP-GAN [Chung et al. 2024]		5.6M	4.1B	6.51	0.712	0.556	0.884	0.785
	<b>DCP-GAN (scaled)</b>				<b>5.68</b>	<b>0.740</b>	0.530	<b>0.981</b>	<b>0.806</b>
	Teacher	small	24.7M	14.9B	4.02	0.769	0.555	1.095	0.854
	GS [Wang et al. 2020]				10.23	0.702	0.430	0.845	0.721
Church	CAGAN [Liu et al. 2021]				9.23	0.685	0.500	0.760	0.722
	StyleKD [Xu et al. 2022]		4.9M	1.3B	7.01	0.707	<b>0.543</b>	<b>0.900</b>	<b>0.783</b>
	DCP-GAN [Chung et al. 2024]				8.17	0.680	0.539	0.787	0.741
	<b>DCP-GAN (scaled)</b>				<b>6.60</b>	<b>0.719</b>	0.532	0.874	0.778
	Teacher	base	30.0M	45.1B	3.97	0.698	0.553	0.849	0.823
	StyleKD [Xu et al. 2022]				5.72	0.703	0.460	0.845	0.777
	DCP-GAN [Chung et al. 2024]		5.6M	4.1B	4.80	0.692	<b>0.505</b>	0.810	0.790
	<b>DCP-GAN (scaled)</b>				<b>4.62</b>	<b>0.716</b>	0.501	<b>0.878</b>	<b>0.812</b>
Church	Teacher	small	24.7M	14.9B	4.37	0.696	0.512	0.815	0.809
	StyleKD [Xu et al. 2022]		1.38B	1.38B	5.99	0.698	0.446	0.848	0.777
	DCP-GAN [Chung et al. 2024]				6.30	0.687	0.435	0.793	0.762
	<b>DCP-GAN (scaled)</b>				<b>5.18</b>	<b>0.719</b>	<b>0.464</b>	<b>0.934</b>	<b>0.805</b>



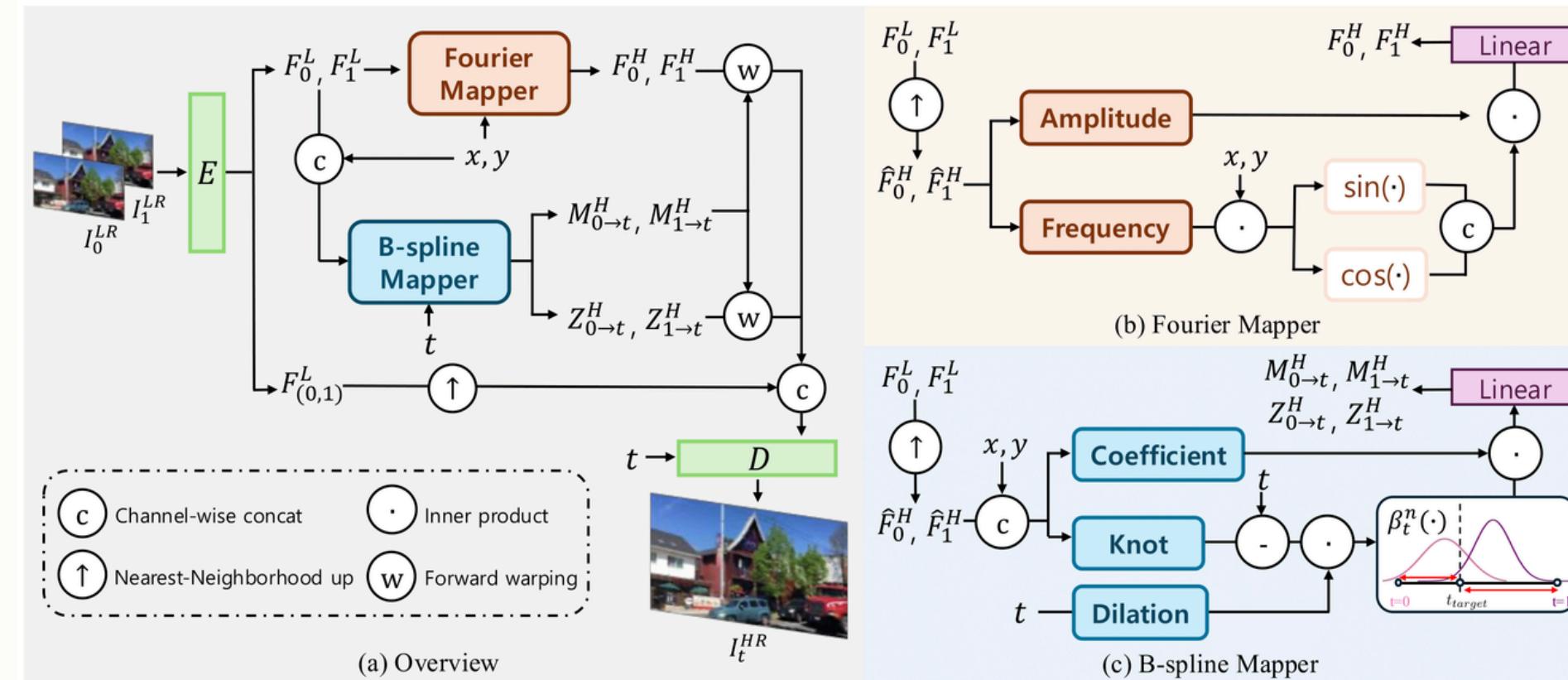
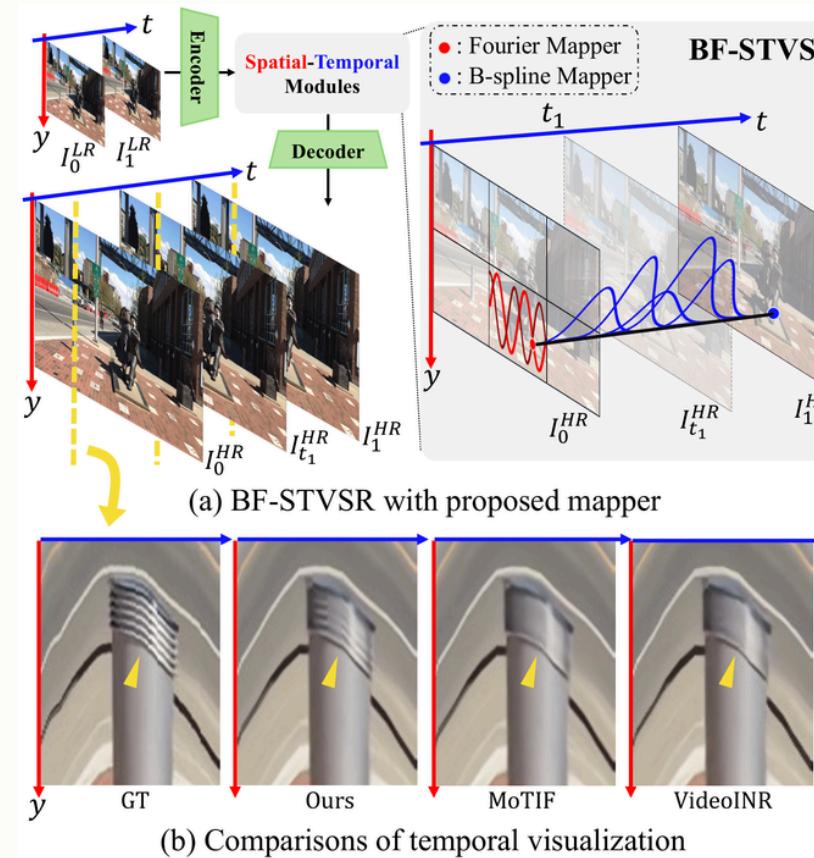
- StyleGAN과 DDPM에서 제안된 가지치기 기법들에서 **Singular Value Scaling**의 성능을 평가하였습니다.
- 대부분의 평가지표에서 성능이 개선**되었으며, **학습 수렴 속도가 크게 향상됨**을 확인할 수 있었습니다.
- 제안된 기법은 **고성능, 고효율의 생성 모델**을 추가 학습 비용 없이, 더욱 적은 학습 시간을 통해 얻을 수 있는 가능성을 제시합니다.

# 시공간 비디오 초해상화 연구

- 고품질의 비디오를 취득하는 것은 **전문화된 장비**를 필요로 하며 **고용량의 저장 매체**를 필요로 합니다.
- 저해상도 저프레임 레이트의 비디오를 **고해상도 고프레임 레이트의 비디오로 복원하는 연구는 사용자에게 차별화된 시각적 경험**을 제공하며, **저장 용량을 절약**하여 경제적으로도 중요도가 높습니다.
- 임의 비율로 비디오를 초해상화하는 **연속 시공간 비디오 초해상화 (C-STVSR)** 연구는 사용자가 원하는 정도의 품질 향상을 제공할 수 있어 큰 주목을 받고 있습니다.
- 본 연구에서 **신호처리 분석을 기반 Positional Encoding** 기법을 제안하여 **C-STVSR 기법의 성능 향상 및 추론 시간을 최적화**하는 것을 목표로 합니다.



# 시공간 비디오 초해상화 연구



## Fourier Mapper

$$A_r \odot \begin{bmatrix} \cos(\pi F_r \delta_r) \\ \sin(\pi F_r \delta_r) \end{bmatrix}$$

## B-spline Mapper

$$p_\psi(z_r, \delta_r, \hat{t}) = c_r \odot \beta^n \left( \frac{\hat{t} - k_r}{d} \right)$$

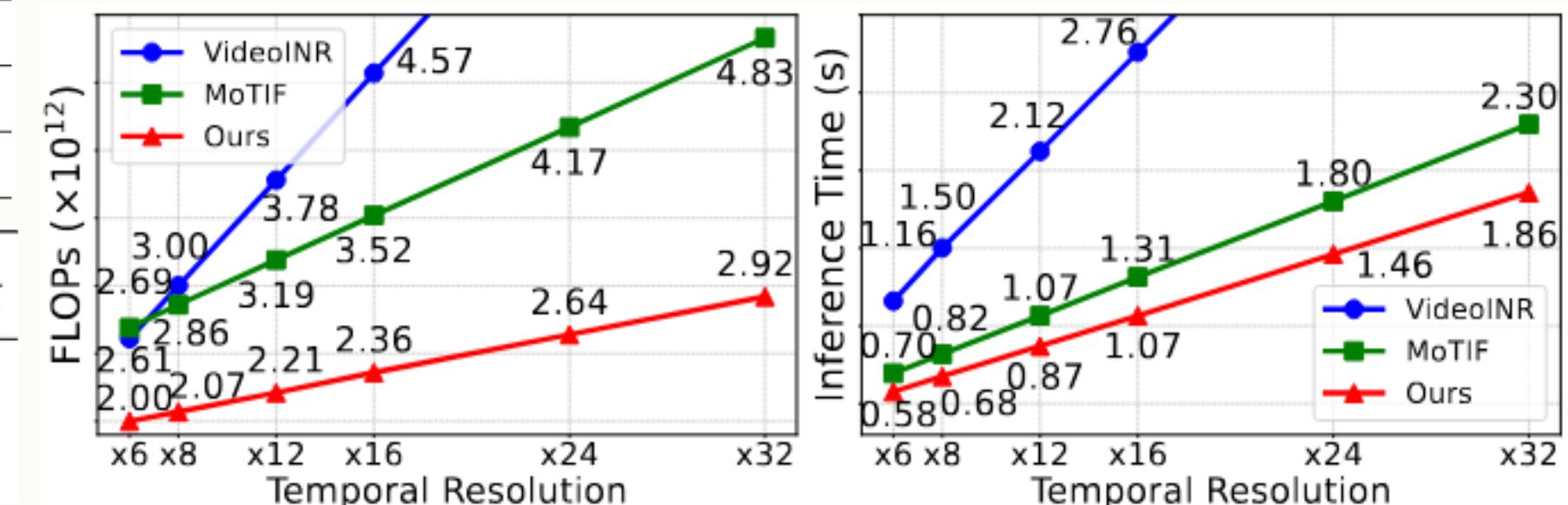
- 기존 C-STVSR 연구들은 비디오의 움직임 및 공간 정보 모델링을 위해 순수 좌표 기반 네트워크(INR)를 사용하였습니다.
- 하지만 이 연구들은 feature에 단순한  $(x, y, t)$  좌표 concat 방식의 **Positional Encoding**은 고주파 정보와 복잡한 움직임을 충분히 포착하지 못해 **예측된 비디오**에 **artifact**를 유발합니다.
- 본 연구에서는 시공간 정보 모델링을 위해 **Fourier & B-spline** 기저함수를 사용한 **Positional Encoding**을 제안하였습니다.
- 공간 정보는 **고주파 신호 인코딩**에 효과적인 **Fourier** 기저함수를 활용해 비디오의 **세밀한 패턴**을 정밀하게 포착하였습니다.
- 시간 정보는 **B-spline** 기저함수를 활용하여 **비디오 움직임의 복잡하지만 연속적인 특성**을 반영하여 움직임을 부드럽게 예측하였습니다.

# 시공간 비디오 초해상화 연구

VFI Method	VSR Method	Vid4	GoPro-Center	GoPro-Average	Adobe-Center	Adobe-Average	Parameters (Millions)
SuperSloMo [14]	Bicubic	22.42 / 0.5645	27.04 / 0.7937	26.06 / 0.7720	26.09 / 0.7435	25.29 / 0.7279	19.8
SuperSloMo [14]	EDVR [39]	23.01 / 0.6136	28.24 / 0.8322	26.30 / 0.7960	27.25 / 0.7972	25.90 / 0.7682	19.8+20.7
SuperSloMo [14]	BasicVSR [3]	23.17 / 0.6159	28.23 / 0.8308	26.36 / 0.7977	27.28 / 0.7961	25.94 / 0.7679	19.8+6.3
QVI [43]	Bicubic	22.11 / 0.5498	26.50 / 0.7791	25.41 / 0.7554	25.57 / 0.7324	24.72 / 0.7114	29.2
QVI [43]	EDVR [39]	23.48 / 0.6547	28.60 / 0.8417	26.64 / 0.7977	27.45 / 0.8087	25.64 / 0.7590	29.2+20.7
QVI [43]	BasicVSR [3]	23.15 / 0.6428	28.55 / 0.8400	26.27 / 0.7955	26.43 / 0.7682	25.20 / 0.7421	29.2+6.3
DAIN [1]	Bicubic	22.57 / 0.5732	26.92 / 0.7911	26.11 / 0.7740	26.01 / 0.7461	25.40 / 0.7321	24.0
DAIN [1]	EDVR [39]	23.48 / 0.6547	28.58 / 0.8417	26.64 / 0.7977	27.45 / 0.8087	25.64 / 0.7590	24.0+20.7
DAIN [1]	BasicVSR [3]	23.43 / 0.6514	28.46 / 0.7966	26.43 / 0.7966	26.23 / 0.7725	25.23 / 0.7725	24.0+6.3
ZoomingSloMo [40]		25.72 / 0.7717	30.69 / 0.8847	- / -	30.26 / 0.8821	- / -	11.10
TMNet [42]		25.96 / 0.7803	30.14 / 0.8696	28.83 / 0.8514	29.41 / 0.8524	28.30 / 0.8354	12.26
VideoINR [7]		25.61 / 0.7709	30.26 / 0.8792	29.41 / 0.8669	29.92 / 0.8746	29.27 / 0.8651	11.31
MoTIF [6]		25.79 / 0.7745	31.04 / 0.8877	30.04 / 0.8773	30.63 / 0.8839	29.82 / 0.8750	12.55
BF-STVSR + $\mathcal{L}_{RAFT}$ (Ours)		25.80 / 0.7754	31.14 / 0.8893	30.20 / 0.8799	30.84 / 0.8877	30.14 / 0.8808	13.47
BF-STVSR (Ours)		25.85 / 0.7772	31.17 / 0.8898	30.22 / 0.8802	30.83 / 0.8880	30.12 / 0.8808	

Method	GoPro				Adobe			
	VFIPS↑	FloLPIPS↓	tOF↓	VMAF↑	VFIPS↑	FloLPIPS↓	tOF↓	VMAF↑
VideoINR	81.13	0.151	0.519	57.96	81.15	0.145	0.574	67.08
MoTIF	81.89	0.156	0.517	59.82	81.61	0.144	0.607	68.40
BF-STVSR + $\mathcal{L}_{RAFT}$ (Ours)	83.26	0.151	0.474	61.09	84.14	0.131	0.488	70.79
BF-STVSR (Ours)	83.01	0.151	0.480	61.06	84.04	0.132	0.498	70.82

Temporal Scale	Spatial Scale	RIFE [13]		EMA-VFI [45]		VideoINR [7]	MoTIF [6]	BF-STVSR + $\mathcal{L}_{RAFT}$ (Ours)	BF-STVSR (Ours)
		LIIF [5]	LTE [16]	LIIF [5]	LTE [16]				
x8	x4	29.14 / 0.8524	29.14 / 0.8524	29.68 / 0.8671	29.68 / 0.8667	29.41 / 0.8669	30.04 / 0.8773	30.20 / 0.8799	30.22 / 0.8802
	x6	30.16 / 0.8738	30.16 / 0.8737	30.64 / 0.8850	30.64 / 0.8848	30.78 / 0.8954	31.56 / 0.9064	31.68 / 0.9082	31.70 / 0.9083
	x12	27.87 / 0.8038	27.86 / 0.8031	28.17 / 0.8126	28.17 / 0.8117	25.56 / 0.7671	29.36 / 0.8505	29.44 / 0.8516	29.45 / 0.8520
x12	x4	24.74 / 0.7019	24.70 / 0.6994	24.85 / 0.7052	24.82 / 0.7028	24.02 / 0.6900	25.81 / 0.7330	25.78 / 0.7284	25.80 / 0.7295
	x6	27.43 / 0.8102	27.42 / 0.8100	27.90 / 0.8263	27.90 / 0.8260	27.32 / 0.8141	27.77 / 0.8230	28.06 / 0.8287	28.07 / 0.8287
	x12	26.19 / 0.7640	26.19 / 0.7636	26.49 / 0.7748	26.49 / 0.7743	24.68 / 0.7358	26.78 / 0.7908	27.06 / 0.7961	27.07 / 0.7963
x16	x4	24.03 / 0.6869	24.00 / 0.6853	24.16 / 0.6918	24.15 / 0.6902	23.70 / 0.6830	24.72 / 0.7108	24.87 / 0.7096	24.88 / 0.7104
	x6	26.08 / 0.7735	26.08 / 0.7733	26.56 / 0.7904	26.56 / 0.7902	25.81 / 0.7739	25.98 / 0.7758	26.40 / 0.7844	26.39 / 0.7840
	x12	25.24 / 0.7394	25.24 / 0.7391	25.54 / 0.7503	25.55 / 0.7499	23.86 / 0.7123	25.34 / 0.7527	25.81 / 0.7621	25.81 / 0.7619



- 제안된 Positional Encoding을 적용한 프레임워크를 다양한 비디오 데이터셋에서 평가하였습니다.
- 정통적 평가지표인 PSNR/SSIM에서 성능이 개선됨은 물론, 다양한 비디오 품질 평가 지표들에서도 우수한 성능을 보였습니다.
- 또한, B-spline 기저함수의 보간 능력 및 CUDA 함수 구현을 통한 연산 최적화를 통해 기존 기법들 대비 효율적인 추론이 가능합니다.
- 제안된 기법을 통해 저품질 비디오를 원하는 정도의 고품질 비디오로 효율적으로 복원할 수 있는 새로운 기술을 제시합니다.

# 김현진

---

Github

<https://github.com/hjinnkim>

Phone

010-4688-1203

Email

khj12203@gmail.com