Code Paper

ICML
International Conference On Machine Learning

Sparse Model Inversion: Efficient Inversion of Vision Transformers for Data-Free Applications

Zixuan Hu[1,2], Yongxian Wei[1], Li Shen[3,4], Zhenyi Wang[5], Lei Li[1], Chun Yuan[1], Dacheng Tao[2]

1 Tsinghua University; 2 Nanyang Technological University; 3 Sun Yat-sen University;
4 JD Explore Academy; 5 University of Maryland, College Park

Motivation

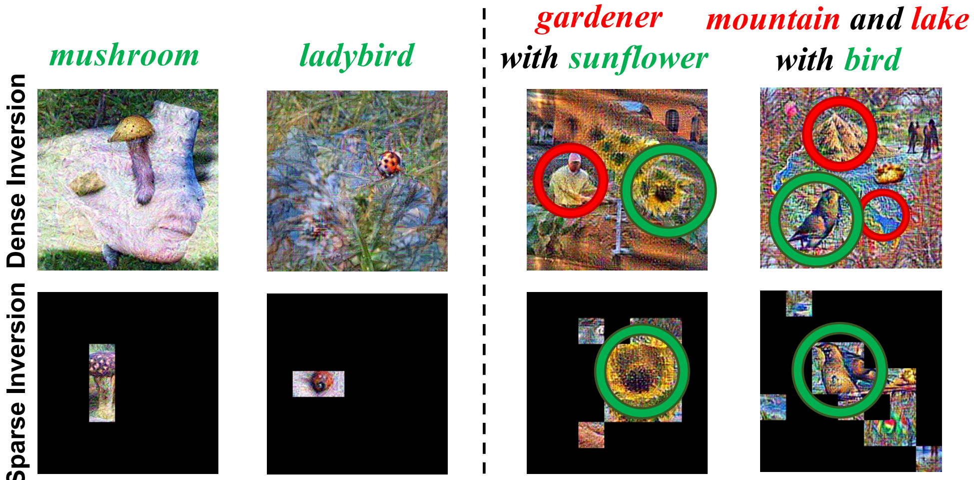
- Limitation of dense model inversion: Extreme inefficiency when inverting high-resolution images from large-scale ViTs

Table 1. Inefficiency of dense inversion (*e.g.*, DeepInversion).

Resolution	**Model**	**Inversion**	**Throughput** (its/s)	**FLOPs** (G)
32 × 32	ResNet18	Dense	77.91	0.11
224 × 224	ResNet18	Dense	10.21	5.47
224 × 224	DeiT/16-Base	Dense	1.79	6475.63

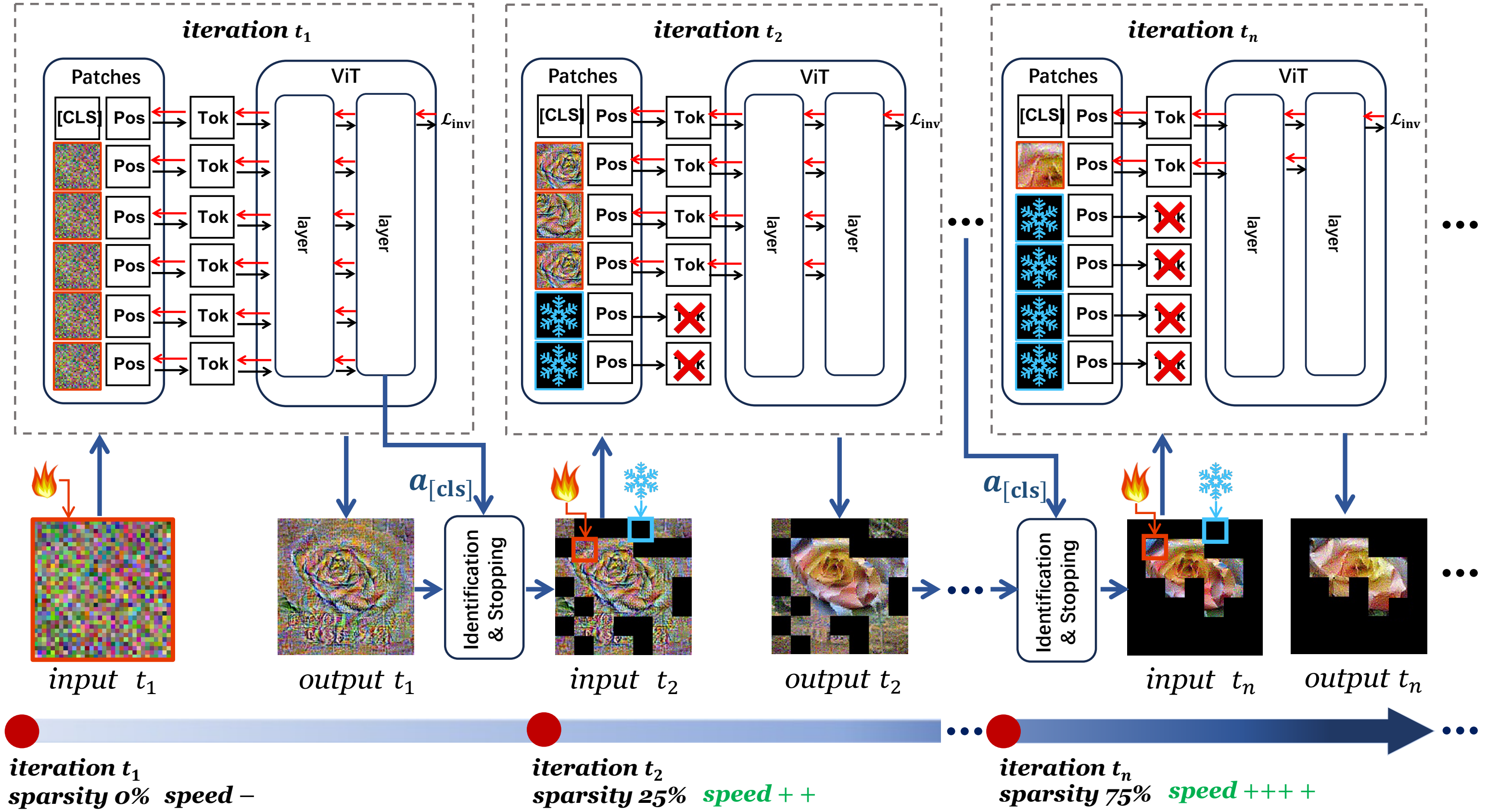
- Cause (a): Redundant inversion of noisy backgrounds

- Cause (b): Unintended inversion of spurious correlations between foregrounds and backgrounds



(a) Redundant inversion of noisy backgrounds
(b) Unintended inversion of spurious correlations

Methodology



➢ **Semantic patch identification**

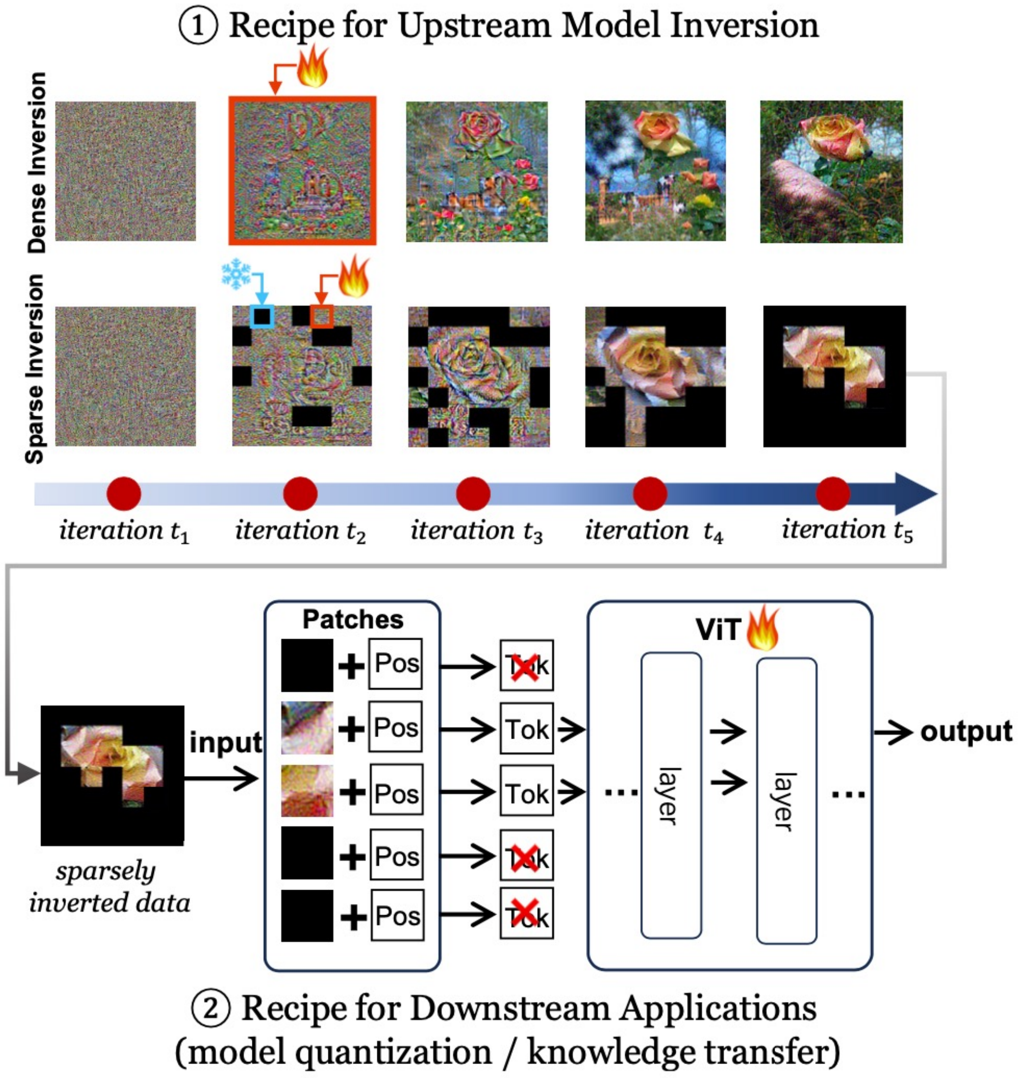
- Identify the importance of each patch via the attention weight $\boldsymbol{a}_{[\mathrm{cls}]}$

$$\boldsymbol{x}_{[\mathtt{CLS}]} = \boldsymbol{a}_{[\mathtt{CLS}]} \cdot \boldsymbol{V}$$

➢ **Early inversion stopping**

- Progressively stop the inversion process of K background patches with smallest $\boldsymbol{a}_{[\mathrm{cls}]}$
- Selectively invert semantic foreground patches with high $\boldsymbol{a}_{[\mathrm{cls}]}$

Methodology



- **Recipe for upstream model inversion and downstream applications**
- Our approach selectively inverts semantic foreground patches while progressively stopping the inversion of uninformative background ones. When utilizing sparsely inverted data for downstream applications, we feed forward only the retained foreground patches.

- **Downstream task I: Data-free model quantization**

$$\theta_{\mathrm{d}} = \left\lfloor \{\mathrm{clip}\left(\theta_{\mathrm{u}}; T_{\min}, T_{\max}\right) - T_{\min}\}/S \right\rceil,$$
$$\text{where } S = \{T_{\max} - T_{\min}\}/\{2^k - 1\}.$$

- **Downstream task II: Data-free knowledge transfer**

$$\theta_{\mathrm{d}} = \min_{\theta_{\mathrm{d}}} \frac{1}{|\mathcal{D}_{\mathrm{u}}|} \sum_{\boldsymbol{x} \in \mathcal{D}_{\mathrm{u}}} \mathrm{KL}\left(f_{\mathrm{u}}\left(\boldsymbol{x}; \boldsymbol{\theta}_{\mathrm{u}}\right)/\tau; f_{\mathrm{d}}\left(\boldsymbol{x}; \boldsymbol{\theta}_{\mathrm{d}}\right)/\tau\right)$$

Sparse Model Inversion VS. Dense Model Inversion

(1) Faster upstream model inversion

- Compared with dense inversion, our approach achieves a range of 2.57 to 3.79-fold speed increase, accompanied by a 74.09%-75.62% reduction in FLOPs and 57.42%-62.98% less GPU memory usage.

Model	Method	Model Inversion (Upstream)			
		Sparsity	Throughput (its/s) ↑	FLOPs (G) ↓	GPU Mem (MB) ↓
DeiT/16 -Tiny	Original	—	—	—	—
	Gaussian Noise	—	—	—	—
	PSAQ-ViT (Dense)	0	0.74	414.20	1648.08
	DeepInversion (Dense)	0	7.33	414.20	1118.69
	DeepInversion (Sparse)	77%	18.82 (×2.57)	107.32(−74.09%)	476.32(−57.42%)
DeiT/16 -Base	Original	—	—	—	—
	Gaussian Noise	—	—	—	—
	PSAQ-ViT (Dense)	0	0.46	6475.63	9327.12
	DeepInversion (Dense)	0	1.19	6475.63	4096.96
	DeepInversion (Sparse)	77%	4.51 (×3.79)	1578.97 (−75.62%)	1516.64 (−62.98%)

Sparse Model Inversion VS. Dense Model Inversion

(2) Better or comparable downstream performance

Downstream task I:
Data-free model quantization

Model	Method	Model Inversion (Upstream)				Quantization (Downstream)			
		Sparsity	Throughput (its/s) ↑	FLOPs (G) ↓	GPU Mem (MB) ↓	Prec.	Top-1	Prec.	Top-1
DeiT/16 -Tiny	Original	—	—	—	—	FP	72.14	FP	72.14
	Gaussian Noise	—	—	—	—	W4/A8	7.80	W8/A8	10.55
	PSAQ-ViT (Dense)	0	0.74	414.20	1648.08	W4/A8	64.97	W8/A8	70.54
	DeepInversion (Dense)	0	7.33	414.20	1118.69	W4/A8	64.28	W8/A8	70.27
	DeepInversion (Sparse)	77%	18.82 (×2.57)	107.32(−74.09%)	476.32(−57.42%)	W4/A8	64.04	W8/A8	70.13
DeiT/16 -Base	Original	—	—	—	—	FP	81.85	FP	81.85
	Gaussian Noise	—	—	—	—	W4/A8	11.09	W8/A8	14.72
	PSAQ-ViT (Dense)	0	0.46	6475.63	9327.12	W4/A8	76.73	W8/A8	78.93
	DeepInversion (Dense)	0	1.19	6475.63	4096.96	W4/A8	75.99	W8/A8	78.58
	DeepInversion (Sparse)	77%	4.51 (×3.79)	1578.97 (−75.62%)	1516.64 (−62.98%)	W4/A8	77.51	W8/A8	79.63

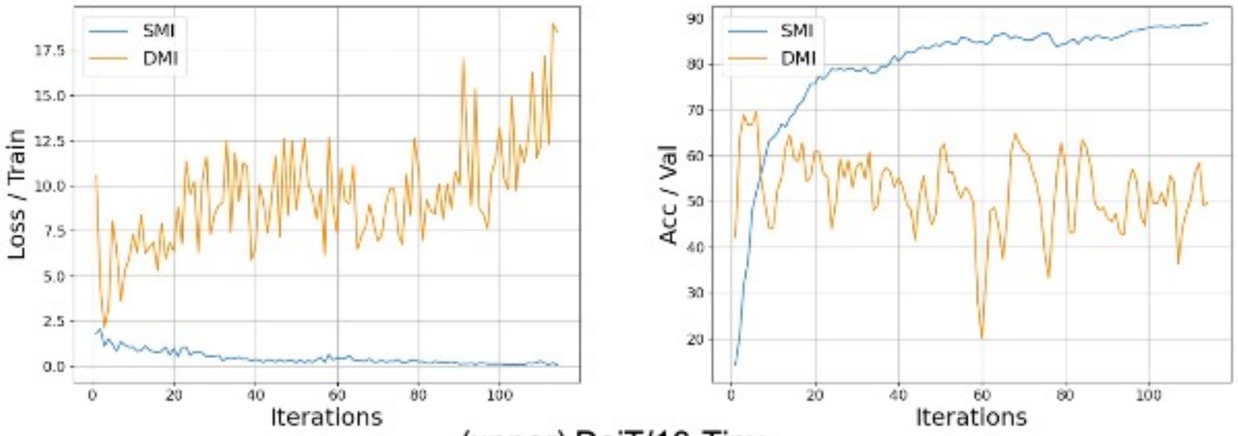
Downstream task II:
Data-free knowledge transfer

Table 4. Knowledge-transfer results on CIFAR10/100 datasets.

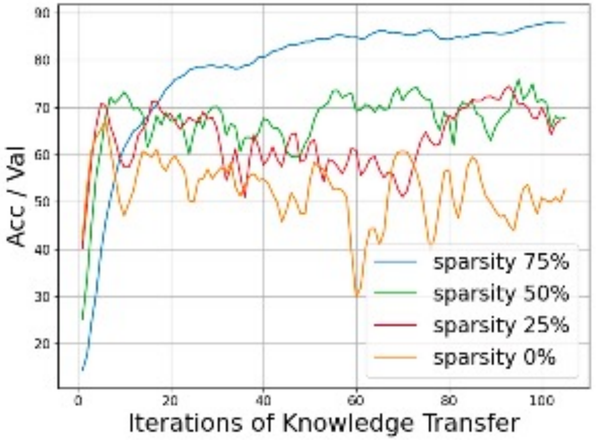
Model	Method	Knowledge Transfer (Downstream)			
		Dataset	Top-1	Dataset	Top-1
DeiT/16 -Tiny	Teacher	CIFAR-10	90.23	CIFAR-100	71.66
	DeepInversion (Dense)	CIFAR-10	69.51	CIFAR-100	70.32
	DeepInversion (Sparse)	CIFAR-10	90.08	CIFAR-100	70.48
DeiT/16 -Base	Teacher	CIFAR-10	95.36	CIFAR-100	79.41
	DeepInversion (Dense)	CIFAR-10	90.02	CIFAR-100	74.88
	DeepInversion (Sparse)	CIFAR-10	95.10	CIFAR-100	74.53

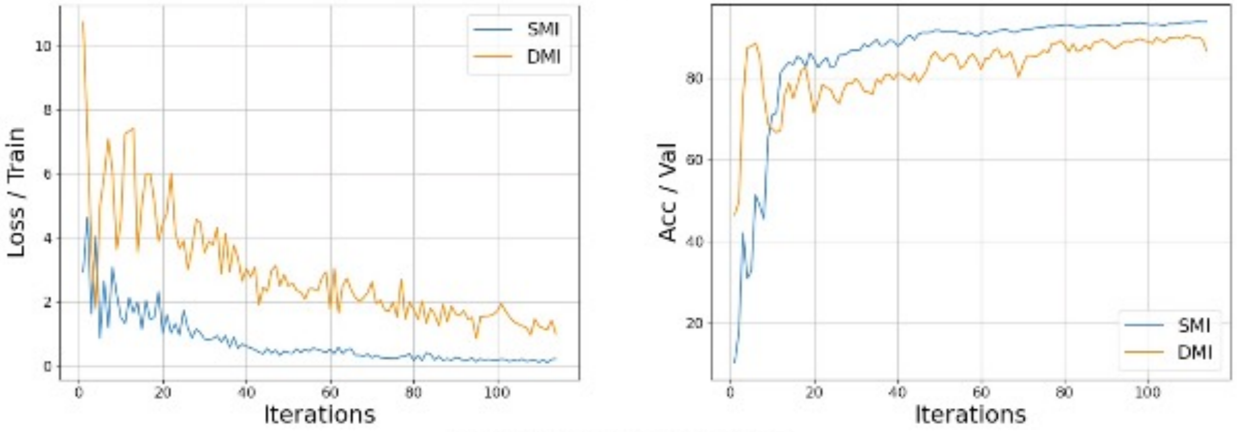
Sparse Model Inversion VS. Dense Model Inversion

(3) Faster and more stable downstream convergence



(upper) DeiT/16-Tiny





(below) DeiT/16-Base

- **Theoretical Analysis**

sparsity ↑

label-relevant patches α ↑ noise level τ ↓

sample complexity $N \downarrow \propto -\alpha$

iteration count $T \downarrow \propto -\alpha^2, 1/(\Theta(1)-\tau)^2$

Visualization

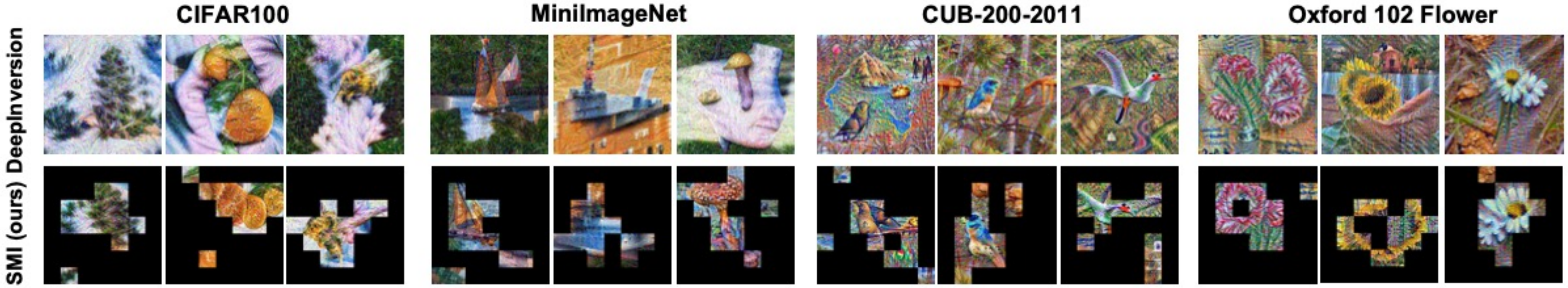


Figure 5. Our inverted images of 224 × 224 pixels from ViT/32-Base encompass a wide range of datasets, from natural images (CIFAR100 and MiniIma- geNet) to more specialized categories (Oxford 102 Flower for various flower species and CUB-200-2011 for bird species).

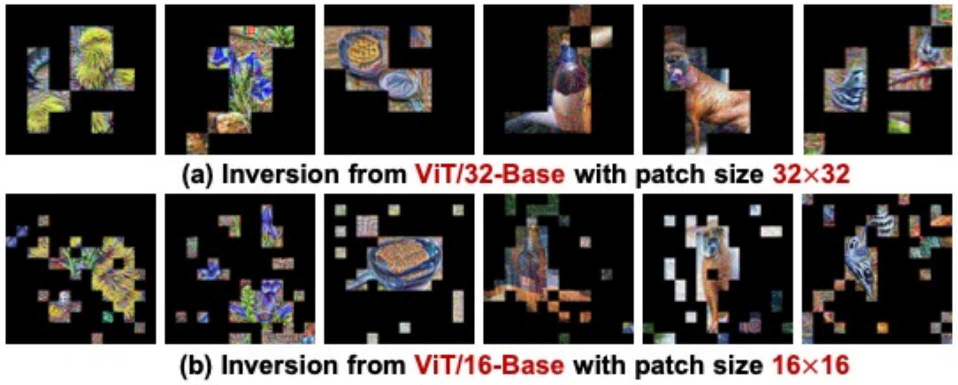


Figure 6. Inversion with different patch-size settings.

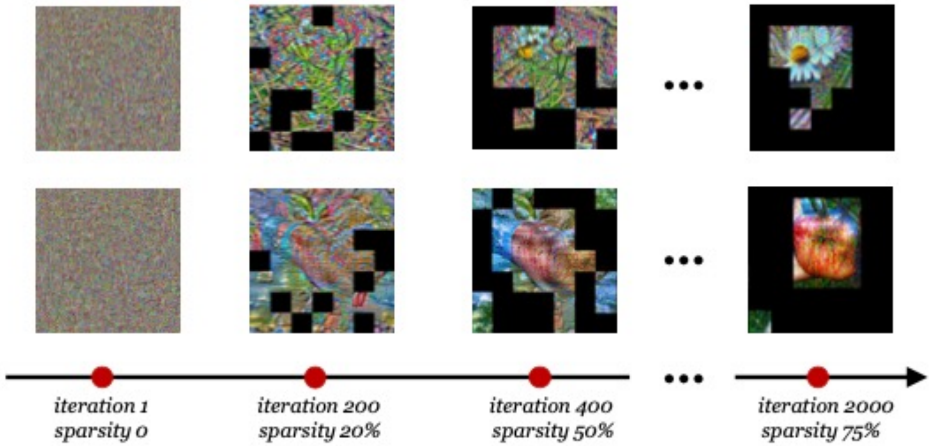


Figure 7. Visualization of the inversion process.

Code Paper

Sparse Model Inversion: Efficient Inversion of Vision Transformers for Data-Free Applications

Zixuan Hu[1,2], Yongxian Wei[1], Li Shen[3,4], Zhenyi Wang[5], Lei Li[1], Chun Yuan[1], Dacheng Tao[2]

1 Tsinghua University; 2 Nanyang Technological University; 3 Sun Yat-sen University;
4 JD Explore Academy; 5 University of Maryland, College Park