Functional Virtual Adversarial Training for Semi-supervised Time Series Classification

Qingyi Pan, Yicheng Li

Tsinghua University

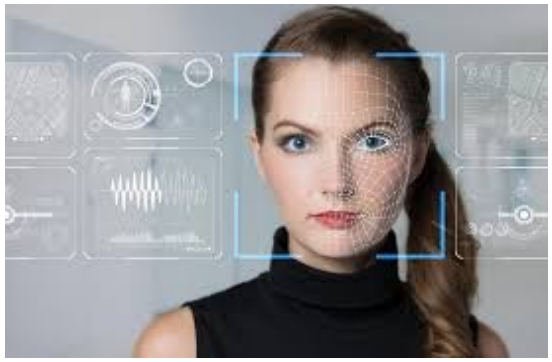
2025.02

Background

- Machine learning are applied in various scenarios
  - Image Classification； Speech Separation； NLP； Time Series；
- The real-world scenarios include

Self-driving car

- ✓ Image recognition
- ✓ Environment perception
- ✓ Autonomous decision



Face Detection

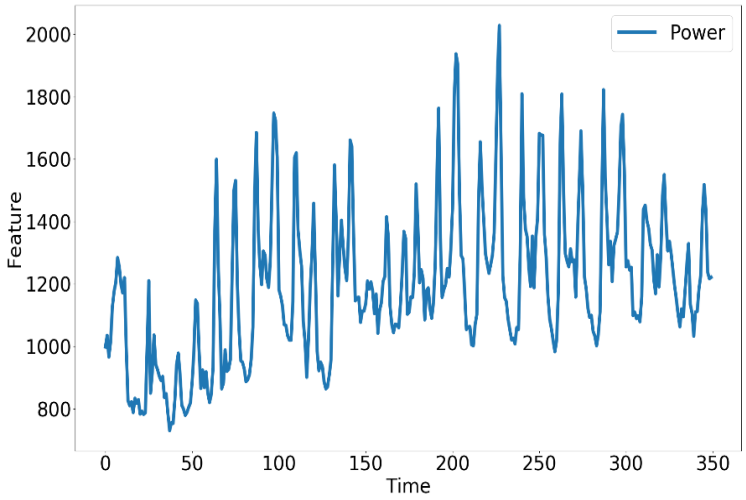
- ✓ Liveness detection
- ✓ Face recognition
- ✓ Facial expression

Game Theory

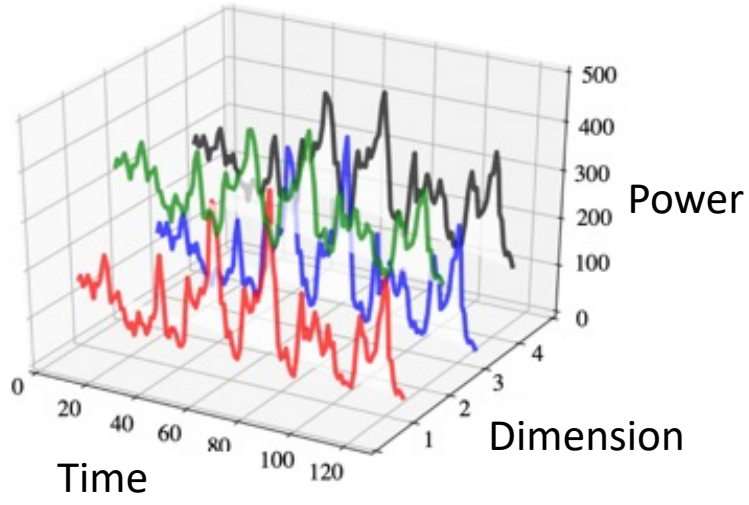
- ✓ Environment perception
- ✓ Decision making
- ✓ Collaborative plan

Background

- Time Series
  - Time series is a sequence of observations indexed in temporal order.
    - $S = [s_1, s_2, \dots, s_t, \dots]. \quad s_t \in R^D$
    - Clustering, Forecasting, Classification, Anomaly Detection



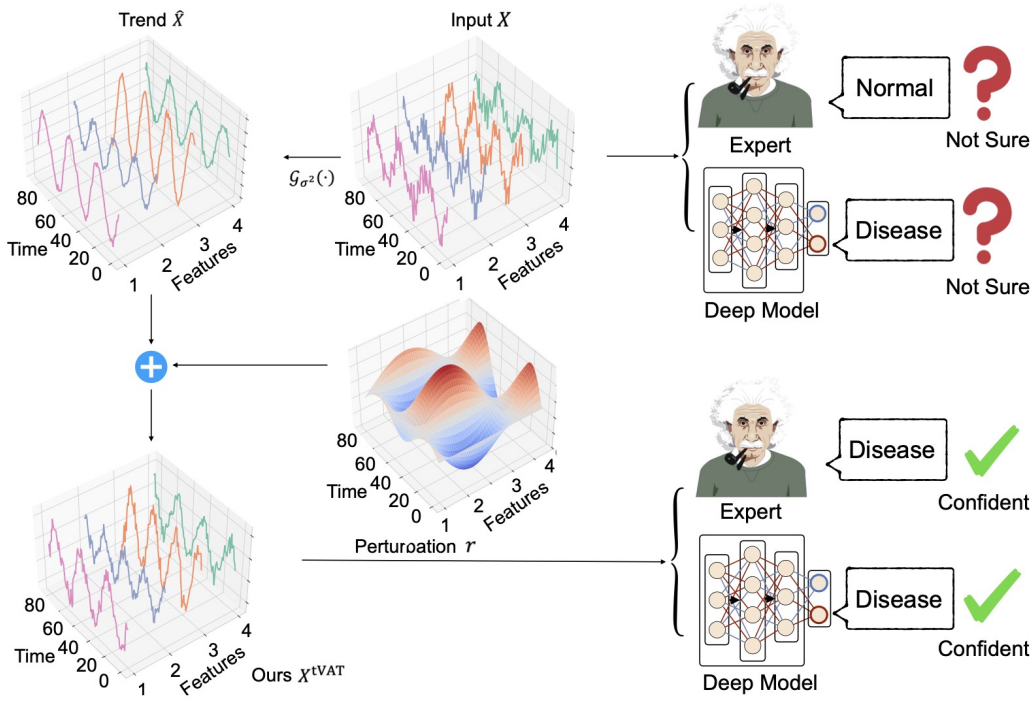
Univariate Time Series



Multivariate Time Series

Background

- Just collect more data?
- Semi-supervised Learning
  - Use unlabeled examples during training
  - Easy to find for time series data.



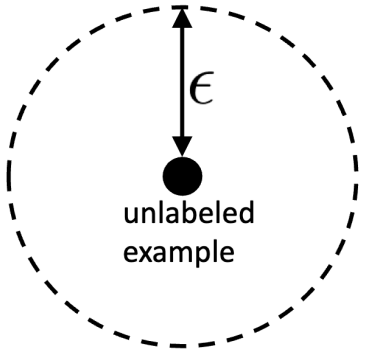
Preliminaries

- Consistency Regularization [Miyato et al., 2018]
  - Add perturbation to the inputs

$$\mathcal{L}_{CE}(p(y|x_j;\theta), p(y|x_j + r;\theta))$$

Soft target Add perturbation

    - Where η is a vector with a random direction and a magnitude ϵ
  - The model should give consistent predictions to nearby samples
    - "Local Distributional Smoothing (LDS)"
    - Perturbation is not chosen randomly (Adversarial)



Motivation

- While for Time Series Data? [Miyato et al., 2018]
  - Create an adversarial example η by maximizing LDS

$$\max \mathcal{L}_{CE}\left(p(y|x_j;\theta), p(y|x_j + r;\theta)\right);\ \ r^* = \epsilon \frac{\nabla_x \mathcal{L}_{CE}}{\|\nabla_x \mathcal{L}_{CE}\|}$$

Soft target Add perturbation

    - where r is not chosen randomly (Adversarial)
    - Introduce abnormal patterns (e.g., Spiky)

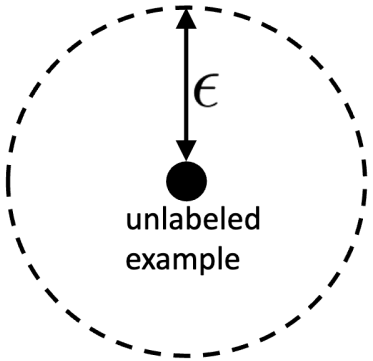
Motivation

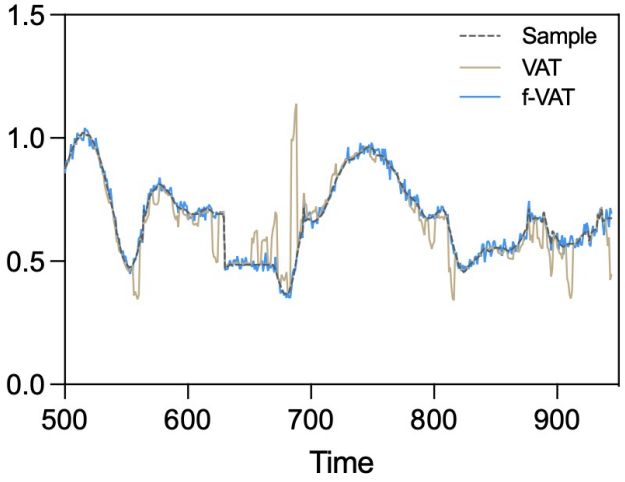
- While for Time Series Data? [Miyato et al., 2018]
  - Create an adversarial example η by maximizing LDS

$$\max \mathcal{L}_{CE}\left(p(y|x_j;\theta), p(y|x_j + r;\theta)\right);\ \ r^* = \epsilon \frac{\nabla_x \mathcal{L}_{CE}}{\|\nabla_x \mathcal{L}_{CE}\|}$$

Soft target — Add perturbation

  - where r is not chosen randomly (Adversarial)
  - Introduce abnormal patterns (e.g., Spiky)





Our Methods

Proposition (Functional linear model)

If $f(X) = \langle \beta, X \rangle_{L^2}$ *with* $\beta \in L^2$, *then*

$$\sup_{\|r\|_E \le \epsilon} \left| f(X+r) - f(X) \right|^2 = \sup_{\|r\|_E \le \epsilon} \langle \beta, r \rangle_{L^2}^2 = \epsilon^2 \|\beta\|_{E^*}^2,$$

with the appropriate dual embedding between E, L^2, *and* E^*.
If $E = \ell_p$ *with dual* ℓ_q $(1/p + 1/q = 1)$, *then*

$$\sup_{\|r\|_p \le \epsilon} \langle g, r \rangle = \epsilon \|g\|_q,$$

Theorem (First-order regime (informal limit))

If f *is Fréchet-differentiable at* X *with gradient* $\nabla f(X) \in L^2$, *then*

$$\lim_{\epsilon \to 0^+} \epsilon^{-2} \sup_{\|r\|_E \le \epsilon} \left| f(X+r) - f(X) \right|^2 = \|\nabla f(X)\|_{E^*}^2.$$

Our Methods

Algorithm 1 Functional Virtual Adversarial Training Step

Input: Data batch $\mathcal{D}, \mathcal{D}^l$, model f_θ, order of the Sobolev norm $s \geq 0$, radius ϵ, adversarial iterations L, learning rate η.
for each sample $X_i \in \mathcal{D}$ **do** ▷ Approximate r_i^*
  Randomly initialize perturbation vector r_i over $\|r_i\|_{H^{-s}} \leq \epsilon$.
  for $\ell = 1 \to L$ **do**
    Gradient ascent $r_i \leftarrow r_i + \eta \nabla_{r_i} \mathrm{LDS}(X_i, r_i; f_\theta)$
    Normalize $r_i \leftarrow \epsilon \frac{r_i}{\|r_i\|_{H^{-s}}}$.
  end for
end for
$\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}(\theta)$, where $\mathcal{L}(\theta) = \mathcal{L}_0(\mathcal{D}^l; f_\theta) + \frac{1}{|\mathcal{D}|} \sum_{X_i \in \mathcal{D}} \mathrm{LDS}(X_i, r_i; f_\theta)$

$$\mathrm{LDS}(X, r; f_\theta) = \left\| f_\theta(X + r) - f_\theta(X) \right\|_2^2, \quad r^*(X) = \arg \max_{\|r\|_{H^{-s}} \leq \epsilon} \mathrm{LDS}(X, r; f_\theta).$$

$$\mathcal{L}(\theta) = \frac{1}{|\mathcal{D}^l|} \sum \mathcal{L}_{\mathrm{CE}}(f_\theta(X), y) + \frac{1}{|\mathcal{D}|} \sum_{X \in \mathcal{D}} \mathrm{LDS}(X, r^*(X); f_\theta).$$

Experiments

- Experimental Setup
  - 30+ UCR/UEA datasets
  - Domestic Futures 50/300/500
  - 0.1/0.2/0.4 label ratios

Dataset	Samples	Length	Dim	Class
CricketX	780	300	1	12
UWave	4478	948	1	8
InsectWing	2200	256	1	11
SelfReg	380	1152	7	2
NATOPS	360	51	24	6
Heartbeat	409	405	61	5

Experiments

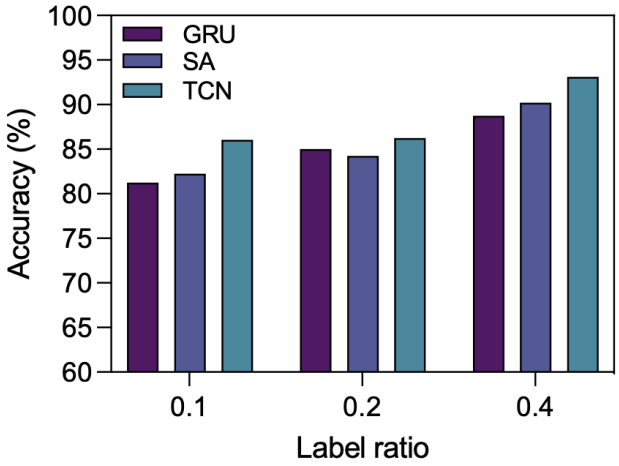
- Empirical Results

Dataset	Ratio	SupL	PI	MTL	SemiTime	TapNet	VAT	f-VAT
CricketX	10%	44.88 ±0.51	38.87 ±2.26	40.94 ±1.97	44.88 ±3.13	39.42 ±0.82	42.85 ±3.97	**49.18** ±1.96
	20%	51.61 ±0.45	44.44 ±2.91	50.12 ±1.22	51.61 ±0.66	51.41 ±0.31	49.14 ±0.50	**57.91** ±3.58
	40%	58.71 ±0.46	53.39 ±2.18	55.10 ±1.12	58.71 ±2.78	58.97 ±0.72	58.63 ±0.50	**68.39** ±2.25
UWave	10%	81.46 ±0.18	81.53 ±0.54	76.35 ±0.56	81.46 ±0.60	82.34 ±0.58	94.41 ±0.09	**94.82** ±0.39
	20%	84.57 ±0.87	81.66 ±0.74	81.77 ±0.94	84.57 ±0.49	86.35 ±0.43	95.53 ±0.31	**96.45** ±0.27
	40%	86.91 ±0.98	86.45 ±1.20	86.91 ±0.68	86.91 ±0.47	89.24 ±0.69	94.76 ±0.54	**97.23** ±0.43
InsectWing	10%	54.96 ±1.25	43.16 ±3.20	50.45 ±1.01	54.96 ±1.61	55.53 ±1.18	55.49 ±1.28	**58.01** ±1.12
	20%	59.01 ±1.13	48.35 ±0.81	56.43 ±0.88	59.01 ±1.56	60.36 ±0.38	61.27 ±0.19	**61.28** ±1.86
	40%	62.38 ±1.39	55.32 ±2.04	60.90 ±0.87	62.38 ±0.76	63.87 ±1.41	63.48 ±0.30	**64.81** ±1.15
SelfReg	10%	46.49 ±2.01	50.44 ±0.76	50.88 ±2.01	49.68 ±2.83	50.87 ±3.31	53.12 ±4.51	**59.31** ±3.06
	20%	52.44 ±3.15	53.94 ±2.63	52.19 ±2.01	52.63 ±1.31	54.39 ±2.74	55.76 ±0.35	**61.60** ±1.13
	40%	51.31 ±3.48	55.69 ±2.74	56.14 ±2.01	49.56 ±1.72	54.38 ±0.76	53.47 ±1.04	**64.44** ±3.13
NATOPS	10%	68.98 ±2.89	75.83 ±4.39	73.91 ±3.73	68.52 ±0.81	70.37 ±7.12	82.38 ±0.96	**86.04** ±1.41
	20%	81.02 ±1.60	82.51 ±1.25	82.41 ±2.89	80.09 ±2.12	77.77 ±1.39	82.81 ±0.52	**86.25** ±1.38
	40%	88.89 ±2.78	88.27 ±1.19	90.27 ±1.39	87.49 ±2.41	82.87 ±2.12	90.15 ±1.60	**93.13** ±0.15
Heartbeat	10%	67.08 ±3.57	72.13 ±1.99	71.61 ±2.47	71.61 ±1.71	72.84 ±1.23	73.86 ±0.59	**76.25** ±1.22
	20%	73.25 ±0.71	72.01 ±0.78	73.66 ±0.71	74.49 ±1.43	73.24 ±1.88	71.59 ±0.13	**76.46** ±1.06
	40%	67.08 ±1.89	73.28 ±1.53	73.61 ±3.07	72.43 ±3.11	73.66 ±0.71	75.00 ±0.11	**77.28** ±0.40

Experiments

- More Empirical Results

Method	10%		20%		40%	
	AvgAcc	AvgRank	AvgAcc	AvgRank	AvgAcc	AvgRank
SupL	35.31	6.67	36.92	7.00	37.15	7.33
PI	53.09	3.93	55.16	4.40	63.60	4.47
MTL	45.19	5.70	45.72	5.87	46.11	6.80
meanTeacher	42.89	5.93	50.94	4.87	63.85	4.13
SemiTime	56.53	3.57	58.93	3.77	69.02	3.13
TapNet	58.67	3.70	60.41	3.40	70.28	3.17
CA-TCC	58.07	3.37	59.84	3.67	63.27	4.27
f-VAT	65.85	1.50	68.87	1.50	76.24	1.53



- Futures Dataset

Futures	Ratio	SupL	PI	MTL	SemiTime	TapNet	CA-TCC	f-VAT
50	10%	40.05 ±1.38	42.87 ±1.45	53.94 ±0.13	55.19 ±0.77	54.62 ±0.54	55.72 ±1.16	**58.64** ±0.53
	20%	45.08 ±1.45	47.26 ±0.73	54.97 ±0.61	56.69 ±0.50	56.93 ±1.28	58.21 ±0.44	**62.09** ±0.26
	40%	50.69 ±3.53	52.42 ±1.21	56.97 ±0.57	57.34 ±0.20	59.75 ±1.09	59.80 ±1.32	**64.78** ±0.45
500	10%	34.23 ±0.24	44.00 ±0.06	39.53 ±1.35	38.86 ±2.04	40.53 ±0.83	39.79 ±1.54	**43.77** ±0.73
	20%	35.38 ±1.06	46.57 ±0.48	45.26 ±0.29	46.04 ±0.12	44.58 ±1.70	45.05 ±1.74	**52.14** ±0.25
	40%	43.85 ±1.49	49.25 ±0.55	47.66 ±1.50	50.65 ±1.05	51.30 ±1.36	54.29 ±0.87	**58.66** ±0.46

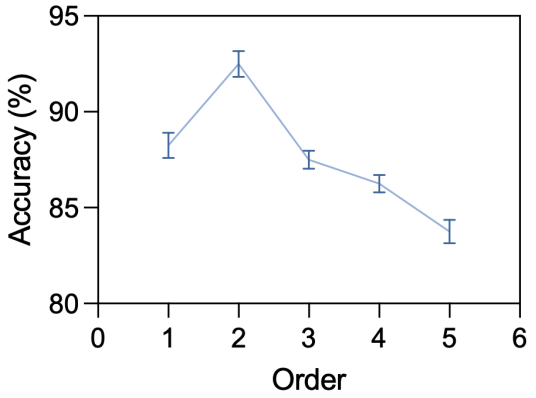
Experiments

- Fully Supervised Performance

Dataset	Hive-COTE	ROCKET	ED	TapNet	ShapeNet	VAT	f-VAT
CricketX	74.10 ±0.03	76.10 ±0.01	62.90 ±0.14	66.20 ±0.25	68.30 ±0.51	68.54 ±1.40	**77.25** ±0.94
UWave	92.10 ±0.02	93.70 ±0.04	88.10 ±0.12	89.40 ±0.69	90.60 ±0.13	92.43 ±0.47	**97.75** ±0.13
InsectWing	62.20 ±0.01	64.70 ±0.01	60.20 ±0.13	67.30 ±0.11	66.30 ±0.02	70.01 ±1.16	**71.70** ±0.55
SelfReg	51.60 ±0.67	51.40 ±0.59	48.30 ±0.12	55.10 ±0.26	57.80 ±0.03	58.75 ±1.25	**60.21** ±0.68
NATOPS	82.80 ±0.32	88.50 ±0.44	85.10 ±0.18	93.90 ±0.01	88.30 ±0.03	87.58 ±1.89	**97.50** ±0.51
Heartbeat	72.20 ±0.52	71.70 ±0.02	61.90 ±0.09	72.10 ±1.43	75.60 ±0.02	76.08 ±0.82	**78.75** ±0.41

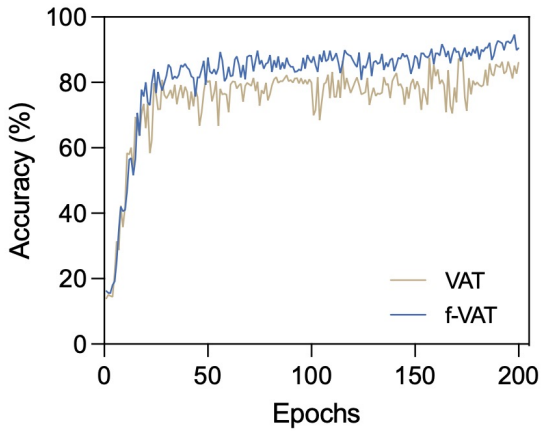
- Altering Order of Sobolev Norm

s	CricketX	UWave	InsectWing	NATOPS	SelfReg
0	58.63 ±0.50	94.76 ±0.54	63.48 ±0.30	90.15 ±1.60	53.47 ±1.04
1	59.91 ±2.32	96.54 ±0.67	66.70 ±0.50	89.58 ±0.12	56.16 ±1.73
2	61.66 ±2.33	97.16 ±0.28	67.08 ±0.86	93.13 ±0.15	58.86 ±0.35
3	60.44 ±0.23	96.82 ±0.16	64.10 ±0.86	90.10 ±0.65	51.39 ±1.21
4	58.22 ±3.39	96.71 ±0.61	66.11 ±0.82	87.51 ±0.52	50.93 ±0.12

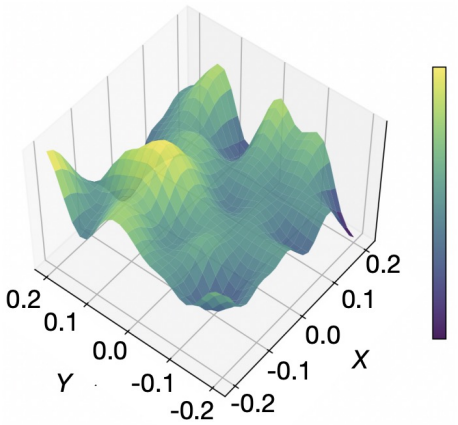


Experiments

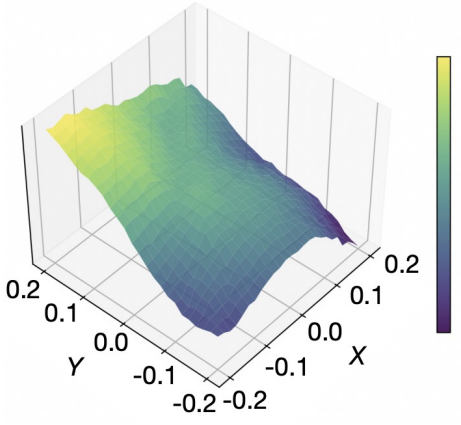
- Visualization of Loss Landscape



(a) Accuracy Curve



(b) VAT



(c) f-VAT

- Runtime Comparison

Method	CricketX	UWave	InsectWing	NATOPS	SelfReg
VAT	15.67	51.66	35.58	28.04	30.68
f-VAT	20.45	62.05	45.92	38.98	44.95
Δ (%)	30.50	20.11	29.06	39.02	46.51

Conclusion

- We propose the framework of functional Virtual Adversarial Training that construct perturbations in various function spaces.
- We theoretically establish the duality between the perturbation norm and gradient sensitivity to generate structured perturbations.
- We propose to use an appropriate Sobolev norm to capture low-frequency trend information and better generalization.

Thanks!