JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

FUDAN UNIVERSITY 1905
復旦大學
FUDAN UNIVERSITY

CAP: Robust Point Cloud Classification via Semantic and Structural Modeling

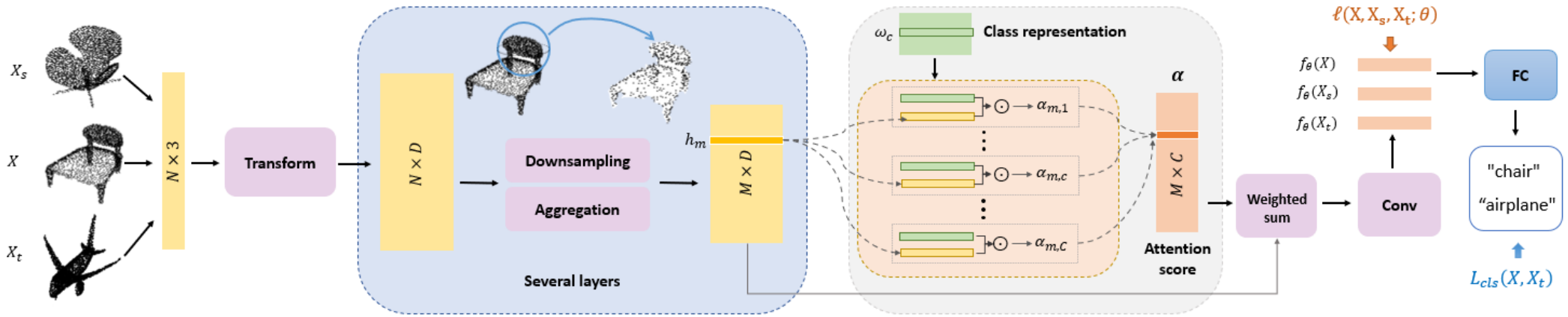
Daizong Ding, Erling Jiang, Yuanmin Huang, Mi Zhang*, Wenxuan Li, Min Yang*
* Corresponding Author

Whitzard-AI Group
School of Computer Science, Fudan University

Poster ID: WED-AM-385

CAP: An Overview

- CAP is a general defense framework for training robust point cloud classification models
- CAP enhances the semantic and structural modeling ability of various existing classifiers
- CAP provides a robustness certification algorithm for potential adversarial attacks



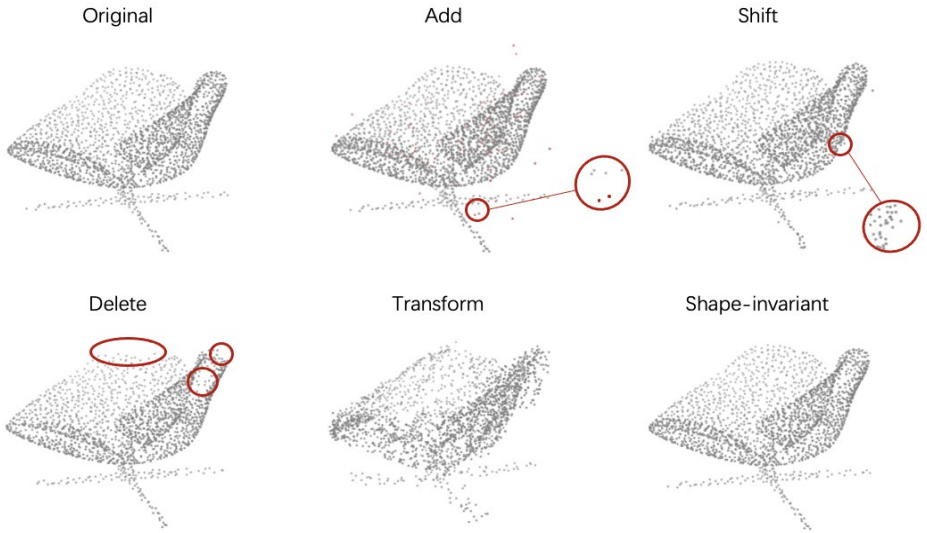
Adversarial Threats for PC Classification

Point cloud classification

- $F: X \mapsto y$, where $X \in \mathbb{R}^{N\times 3}, y \in \{1, \cdots, C\}$

Various emerging adversarial attacks

- Distinct perturbation types
- Diverse adversarial example properties



Poor generalizability of existing defenses

- Adversarial training-based: can only be effective on seen attacks
- Recovery-based: can be evaded by shape-invariant attacks

Our Proposed CAP

Motivation

- Adversarial example preserves semantic and structural information
- Existing classifiers pay attention to limited segments or local features

Contrastive and Attentional Point cloud learning

- Attention-based feature pooling
- Dynamic contrastive learning

Certified robustness

- An algorithm to theoretically certify the robustness of a classifier
- Based on manifold learning and extreme value theory

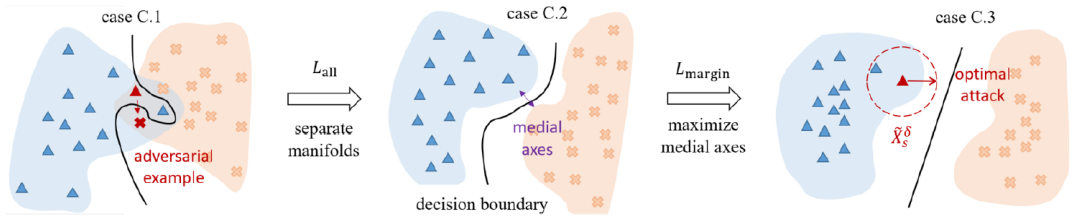
The Comprehensive Framework

Attention-based feature pooling

- Replace the max-pooling of current classifiers with class-wise attention
- Compute the attention score for point i w.r.t. class c as $\alpha_{i,c} = \frac{\exp(h_i^T \omega_c/\tau)}{\sum_{c'} \exp(h_i^T \omega_{c'}/\tau)}$
- Aggregate point features, then extract global feature by $\hat{h}_c = \sum_{i=1}^{N} \alpha_{i,c} h_i$, $z = \mathrm{conv}([\hat{h}_i, \cdots, \hat{h}_C], W^h)$

Dynamic contrastive learning

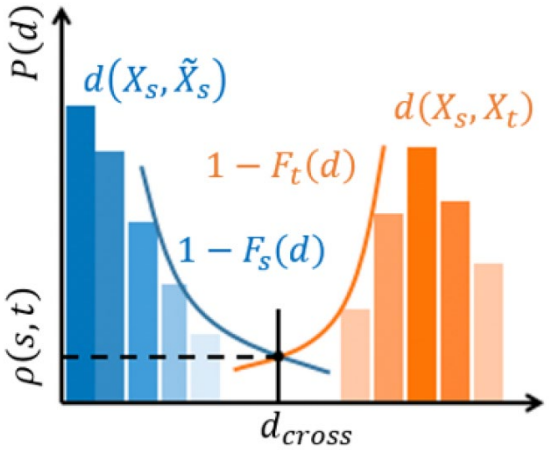
- Triplet loss: $\ell(X, X_s, X_t; \theta) = \max(d(X, X_s) - d(X, X_t) + \epsilon, 0)$
- Margin loss: $L_{margin} = \max_{X_s, X_t \in B(X)} \ell(X, X_s, X_t; \theta)$
- Dynamic learning paradigm
  - $\min_{\theta} L_{all}(\theta; X) + \exp[\epsilon - \bar{L}_{all}(\theta; X)] \cdot L_{margin}(\theta; X)$



Robustness Certification

Manifold learning

- Consider samples from class s lie on manifold $\mathcal{M}_s$
- Sample distance between $X_i, X_j \in \mathcal{M}_s$: $d(X_i, X_j) = \left|\left|z_i - z_j\right|\right|_2$



Robustness certification

- Maximal feature bias within class: $m(\mathcal{M}^s; \delta) = \sup_{X_s \in \mathcal{M}, ||\eta||_p \leq \delta} d(X_s, \tilde{X}_s)$
- Medial axes feature distance between classes: $r(\mathcal{M}^s, \mathcal{M}^t) = \inf_{X_i \in \mathcal{M}^s, X_j \in \mathcal{M}^t} d(X_i, X_j)$
- EVT-based estimation of the CDF of distances: $P(d_* > d) \approx \frac{v}{V} \cdot \left[1 + \gamma_* \cdot \frac{d_*^{(v)} - d}{\beta_*}\right]^{-\frac{1}{\gamma_*}}, d > d_*^{(v)}$
- Certified probability of $m > r$: $\rho(s,t) = P(d_s > d_{cross}) = P(d_t < d_{cross})$

Experimental Setting

Dataset

- ModelNet40, ShapeNet

Table 1. Dataset description.

Dataset	Train size	Test size	# of classes
ModelNet40 [29]	9,843	2,468	40
ShapeNet [2]	35,708	15,419	55

Base model

- PointNet, DGCNN, PointCNN

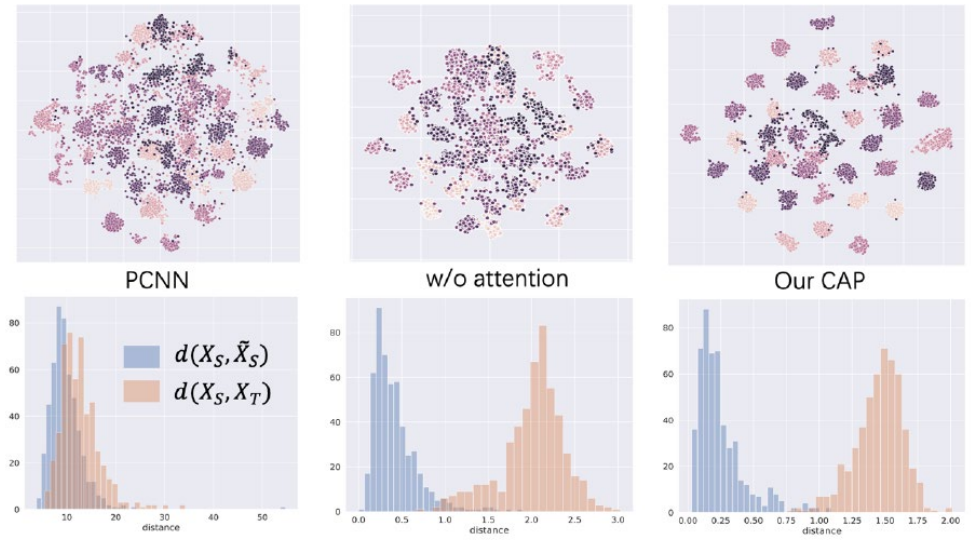
Nine attack baselines

- Normal attacks: Minimal, Smooth, IFGM, PGD, Gen3D-Add, Gen3D-Pert
- Shape-invariant attacks: KNN, GeoA3, SI

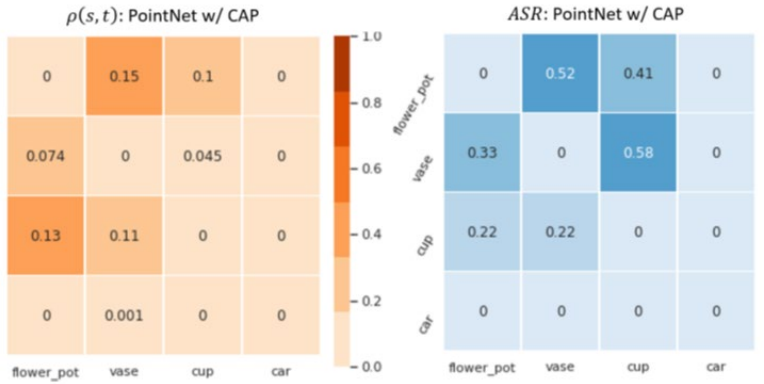
Seven defense baselines

- Adversarial training (AT)-based: AT, AT-PGD, Ensemble AT, PAGN, GvG
- Recovery-based: SOR, DUP-Net

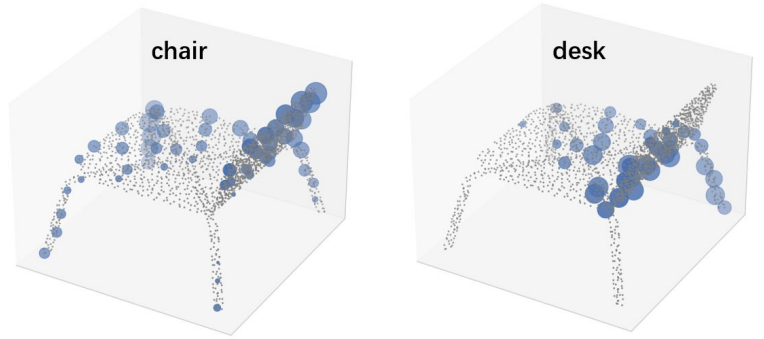
Experimental Results



CAP separates learned features



CAP provides certified robustness consistent with empirical results



CAP helps classifiers pay attention to different parts of a point cloud

Conclusion

Conclusion

- We point out that adversarial attacks preserve key characteristics of original samples
- We develop CAP to enhance the modeling of semantic and structural information
  - Attention-based feature pooling: automatically focuses on important parts
  - Dynamic contrastive learning: coarse-to-fine training separates features
- We provide robustness certification against potential adversarial attacks

Future work

- Validating CAP on voxel-based and transformer-based classifiers
- Extending CAP to other point cloud applications, e.g., 3D object detection / segmentation
- Apply CAP to various domains other than point cloud tasks

Thank you for listening!

If you have any questions, please contact us.

Website of Whitzard-AI Group: https://whitzard-ai.github.io/