Joint Supervised and Self-Supervised Learning for 3D Real World Challenges

Antonio **ALLIEGRO**, Davide **BOSCAINI**, Tatiana **TOMMASI**

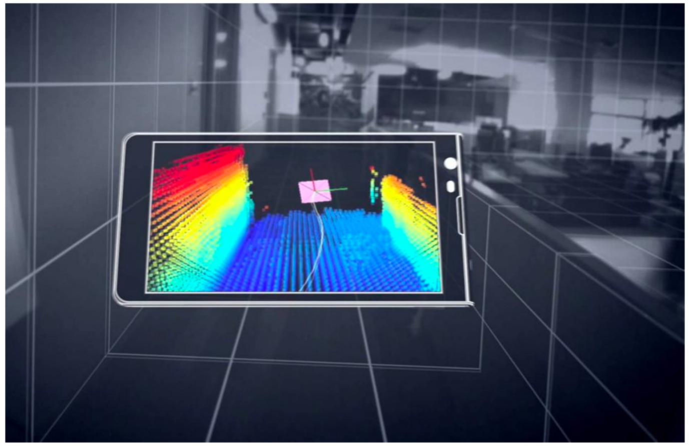
POLITECNICO DI TORINO 1859 1906

FONDAZIONE BRUNO KESSLER

ISTITUTO ITALIANO DI TECNOLOGIA

EMERGING 3D APPLICATION

ROBOTICS



AUGMENTED REALITY

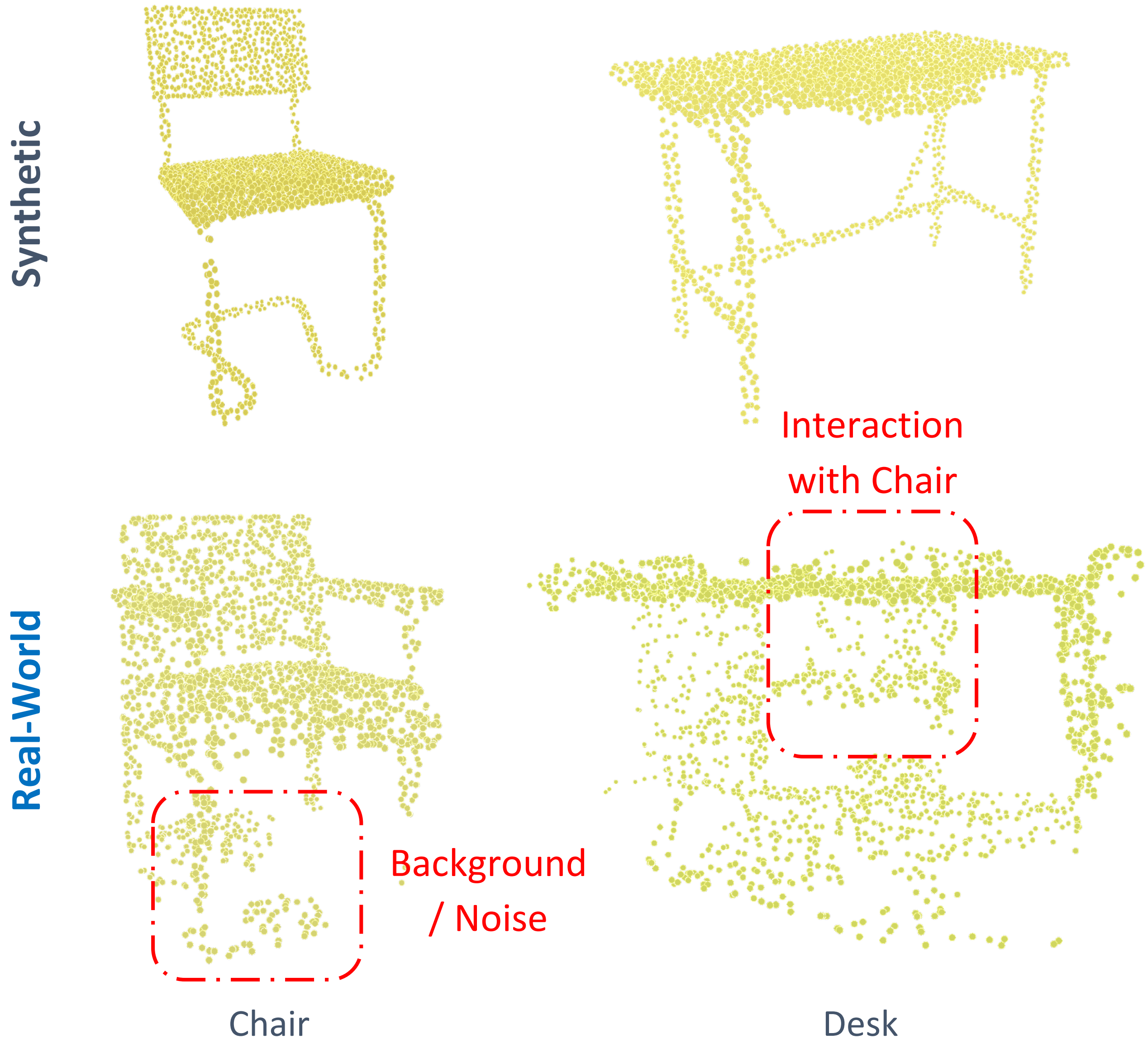
AUTONOMOUS DRIVING

SOURCES: Google Tango, Boston Dynamics

3D REAL-WORLD CHALLENGES

- Are we done with Point Cloud Obj. Classification?
- **What about Real-World data?**
- Deep Learning methods require huge quantity of annotated data
- Real-World datasets are small (ScanObject ~3K samples)
- Can we **build robust models with limited supervised data**?
- Can we **transfer knowledge between Synthetic and Real-World domains**?

SYNTHETIC VS REAL-WORLD



SYNTHETIC

- **ModelNet40** [*Wu et al. in CVPR 2015*], **ShapeNet** [*Chang et al., arXiv 2015*]
- Point Clouds sampled from CAD Models
- Huge availability of annotated samples
- Clean Geometry, No Background

REAL-WORLD

- **ScanObjectNN** [*Uy et al. in ICCV 2019*]
- Background, Partiality, Object Clutter ..
- **No sizable quantities of annotated data!**

SYNTHETIC VS REAL-WORLD

$$\mathbf{P}_{Synth}(X, Y)$$

$$\neq$$

$$\mathbf{P}_{Real}(X, Y)$$

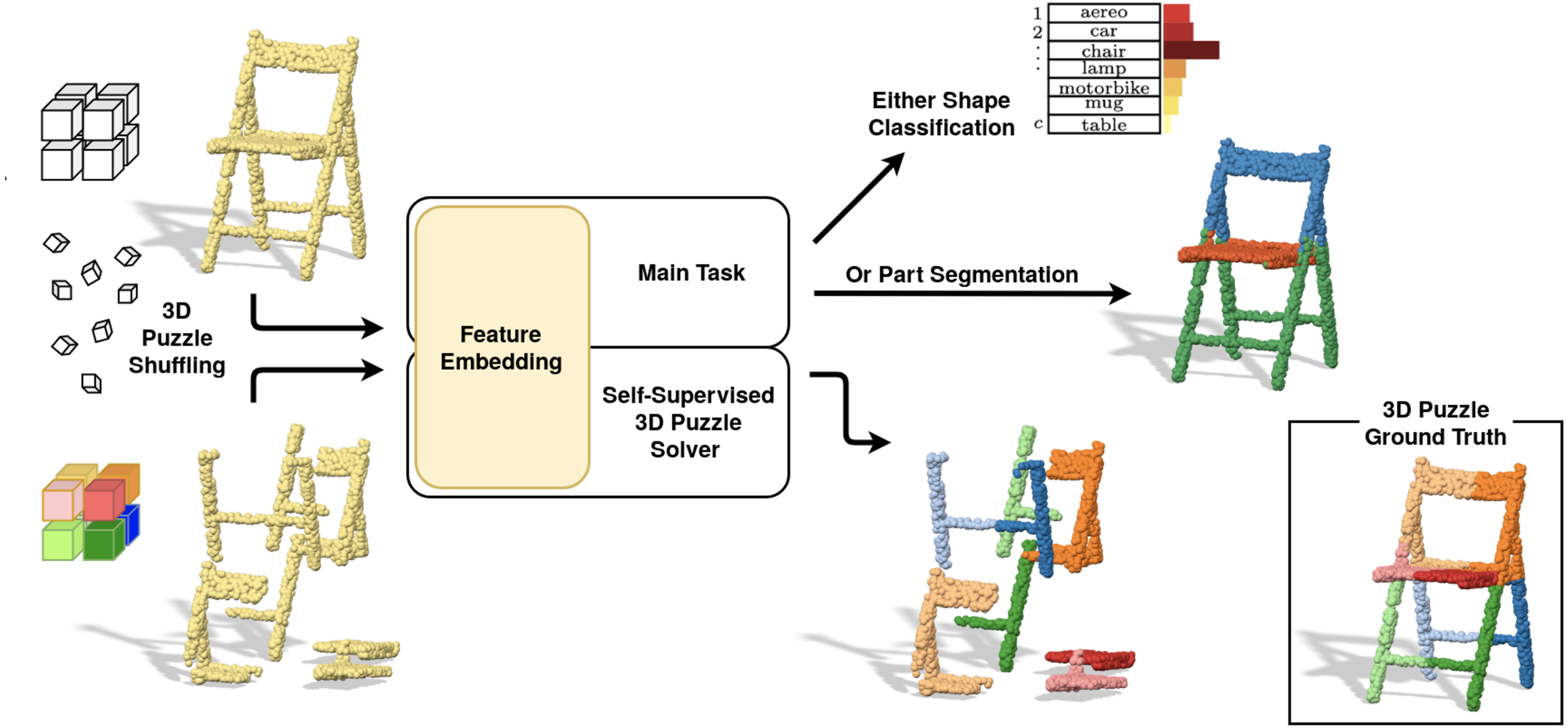
DOMAIN SHIFT

- Synthetic and Real-World Point Cloud data do not share the same distribution
- **Models trained on Synthetic (*Source*) poorly perform on Real-World (*Target*)**

OUR REAL-WORLD RECIPE

- **Multi-Task** approach jointly **combining Supervised and Self-Supervised learning**
- Domain Adaptation and Generalization through self-supervision
- Robustness to limited annotated data (Few-Shot and Semi-Supervised settings)

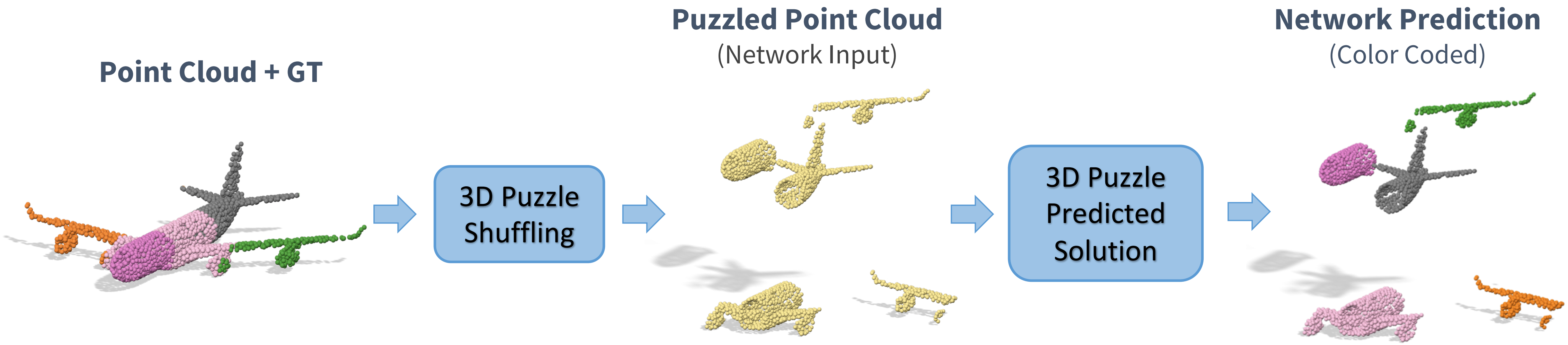
OUR MULTI-TASK



$$\mathcal{L}_{Tot} = \mathcal{L}_m + \alpha \mathcal{L}_p$$

3D PUZZLE

- For each point our network predicts its original membership voxel
- Object understanding is fundamental to solve this SS task:
  - *(learning that..)* **Objects are made of parts**
  - **Spatial Relationship between Objects' parts**

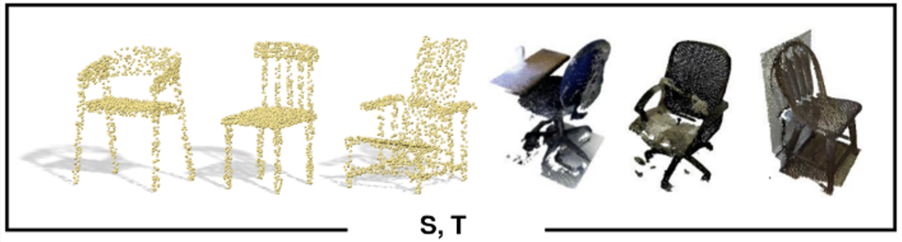


OBJ. CLASSIFICATION - DG / DA

- **DOMAIN GENERALIZATION (DG)**
  The whole set of annotated samples from Source (S) is available at training time.
  We test on the unseen Target collection (T)

- **DOMAIN ADAPTATION (DA)**
  Aims to mitigate the domain shift between Source (S) and Target (T) domain by leveraging on unlabeled Target samples

	training time		test time	
	labeled source S	unlabeled target T	test S	target T
Domain Generalization				
Domain Adaptation				



OBJ. CLASSIFICATION - DG / DA

Classification - Domain Generalization and Adaptation							
Method		ModelNet40 →					PB_T50_RS_BG → ModelNet40
		OBJ_ONLY	OBJ_BG	PB_T50_RS	PB_T50_RS_BG	AVG	
PointDAN [Qin et al.]		56.42	44.84	48.99	34.39	46.16	54.66
PN	Baseline	54.74	43.58	44.96	34.25	44.38	47.43
	Our DG	54.53	49.68	45.22	36.28	46.43	39.30
	Our DA	58.53	47.58	46.70	35.85	47.16	51.54
PN++	Baseline	52.49	44.00	44.83	34.29	43.90	47.66
	Our DG	57.47	52.42	52.84	38.65	50.34	52.88
	Our DA	60.4	53.89	54.66	39.63	**52.14**	**56.07**
3DmFV [Ben-Shabat et al.]		30.90	24.00	24.90	16.40	24.05	51.50
PointCNN [Li et al.]		32.20	29.50	24.60	19.20	26.37	49.20
SpiderCNN [Xu et al.]		44.20	42.10	30.90	22.20	34.85	46.60
DGCNN [Wang et al.]		49.30	46.70	36.80	27.20	40.00	54.70

Other DA technique

Our SOTA Multi-Task

Previous SOTA ARCHs

Synthetic → Real-World

Real-World → Synthetic

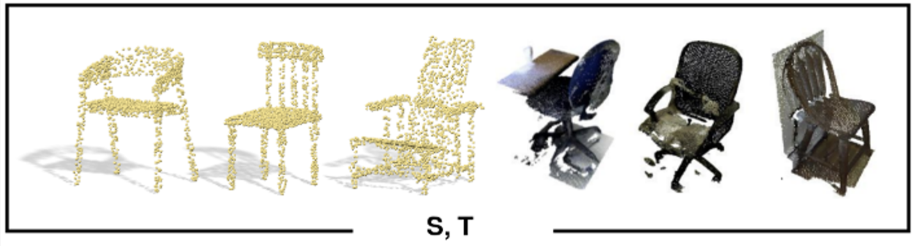
PART SEGMENTATION - FS / SS

We investigate the case of **limited annotated training data** in two settings:

- **FEW SHOT (FS)**
  We reduce at different percentage scales the cardinality of Source Domain training data (S) and evaluate on the whole Source Test (T)

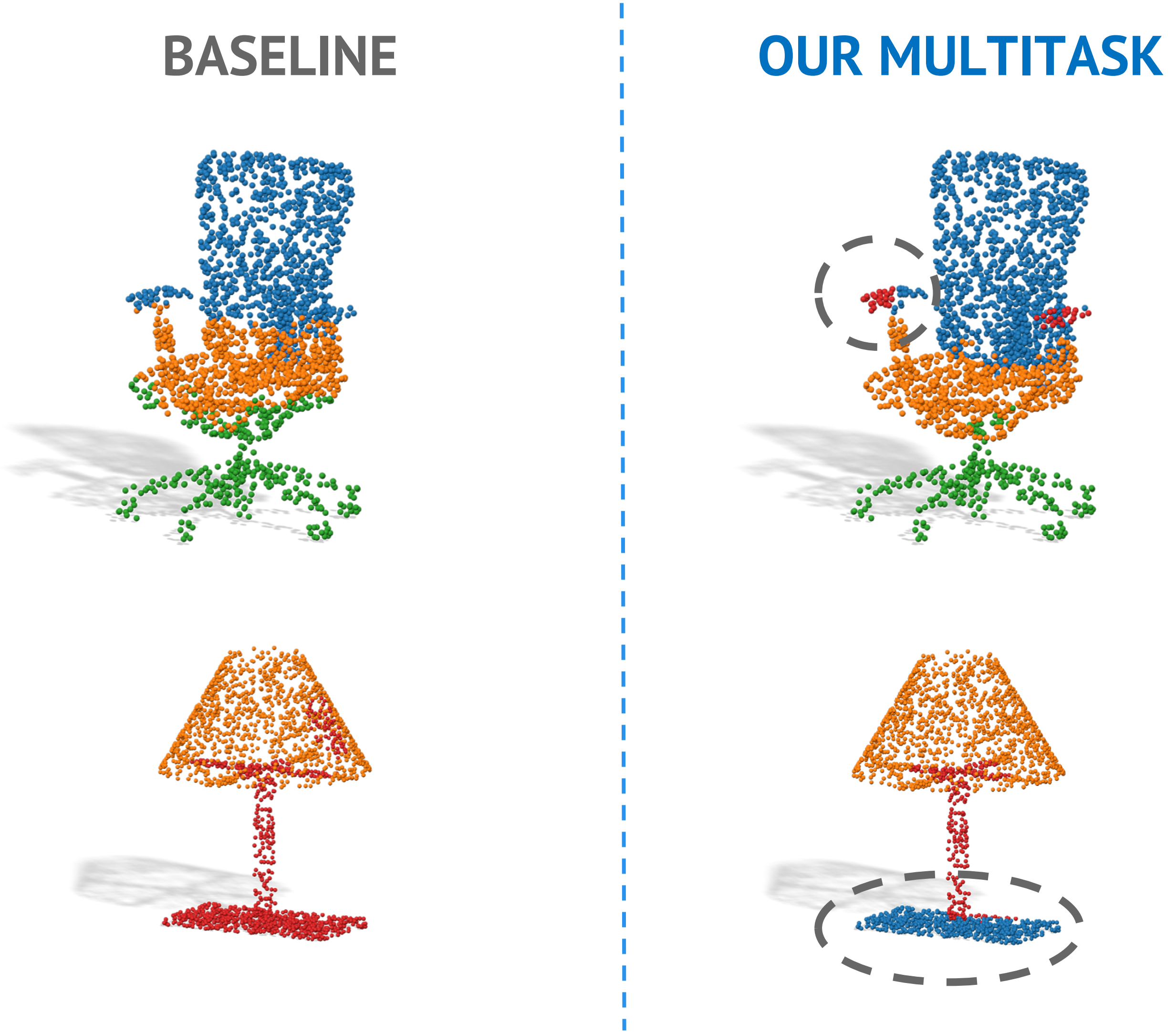
- **SEMI-SUPERVISED (SS)**
  analogous to the Few Shot case but the percentage of samples which is not included in Source Train (S) can be still exploited as unlabeled data

	training time		test time
	labeled source S	unlabeled source S	test S
Few Shot			
Semi-Supervised		Unlabeled	



PART SEGMENTATION - FS / SS

Part Segmentation with limited annotated data		
Method	1%	5%
SO-Net [Li et al.]	64.0	69.0
PointCapsNet [Zhao et al.]	67.0	70.0
CCD [Hassani et al.]	68.2	**77.7**
our baseline	64.52	75.75
our multitask FS	64.49	75.07
our multitask SS	**71.95**	77.42



CONCLUSIONS

Our **multi-task** end-to-end learning framework **combining supervised and self-supervised learning** achieves

- SOTA performances for DA and DG:
  Synthetic ➔ Real-World scenario (and vice-versa)
- Robustness to Limited Annotated Data

We see this work as a first exciting step towards a new family of methods better able to generalize and adapt to novel testing conditions for 3D point clouds.

Thank You!