Generative Zero-Shot Learning for Semantic Segmentation of 3D Point Clouds

Poster presentation - 3DV 2021

Bjoern Michele, Alexandre Boulch, Gilles Puy, Maxime Bucher, Renaud Marlet

SMART TECHNOLOGY FOR SMARTER MOBILITY

Zero-Shot learning

Training

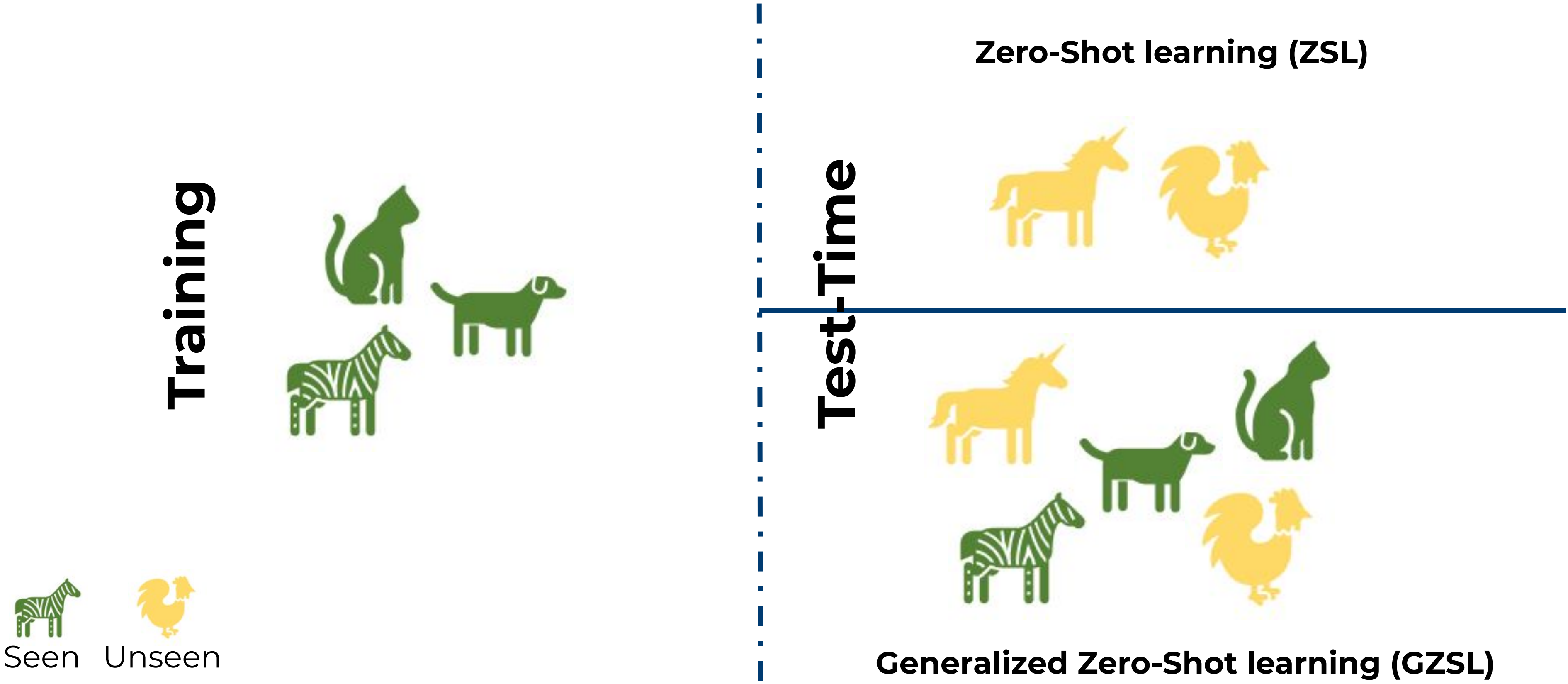
Zero-Shot learning (ZSL)

Test-Time

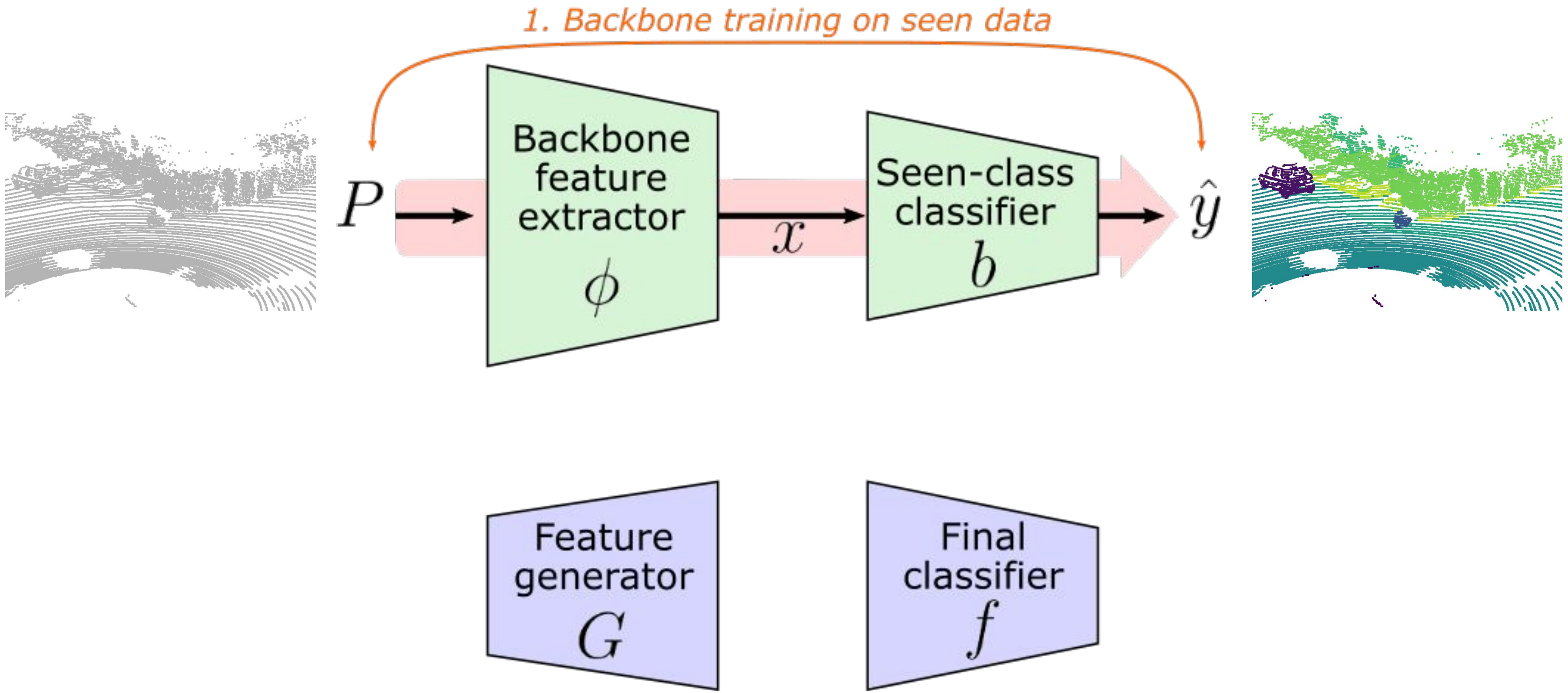
Seen

Unseen

Zero-Shot learning

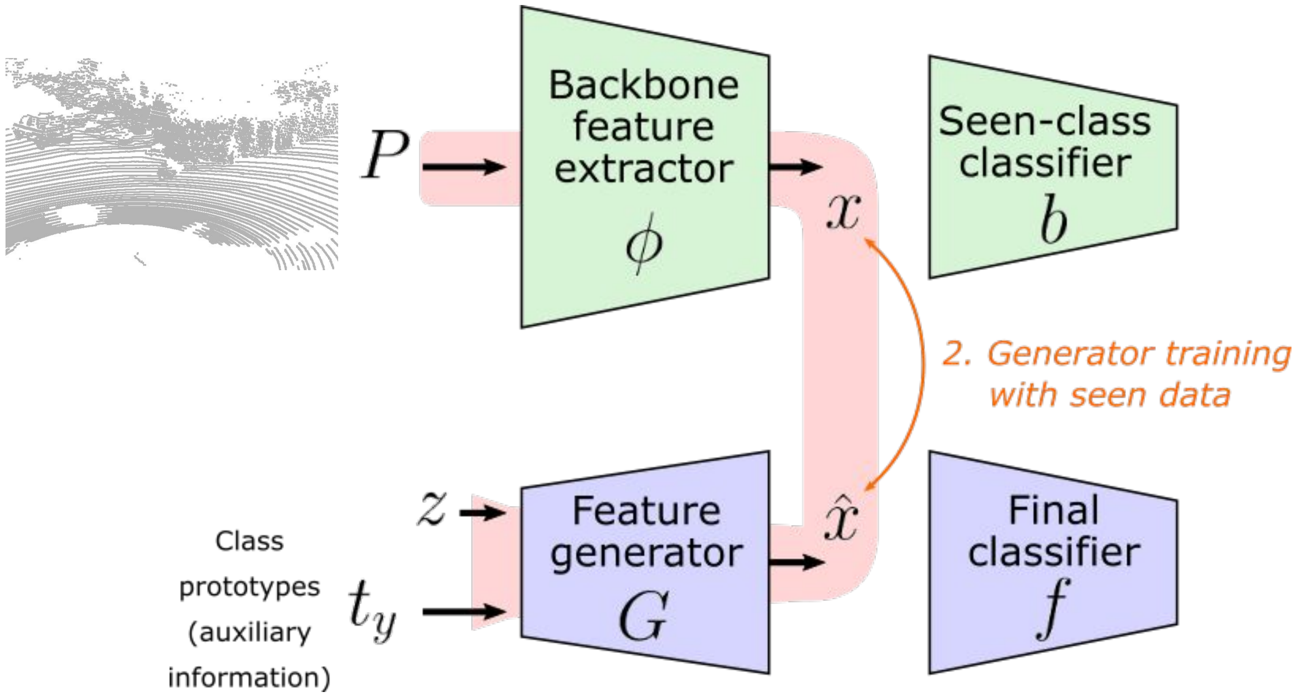


Method[1)]



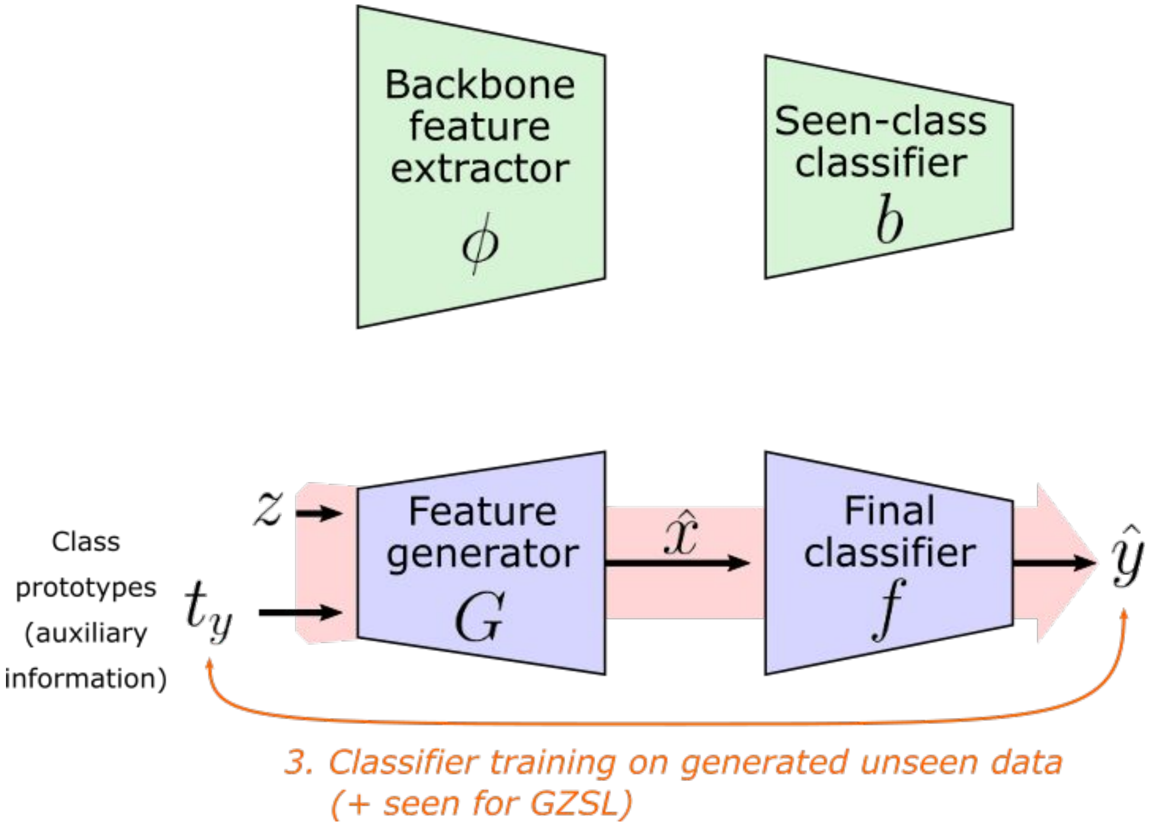
1) Adapted from:
Bucher et al. Generating visual representations for zero-shot classification. ICCV, 2017.
Bucher et al. Zero-shot semantic segmentation. NeurIPS, 2019.

Method



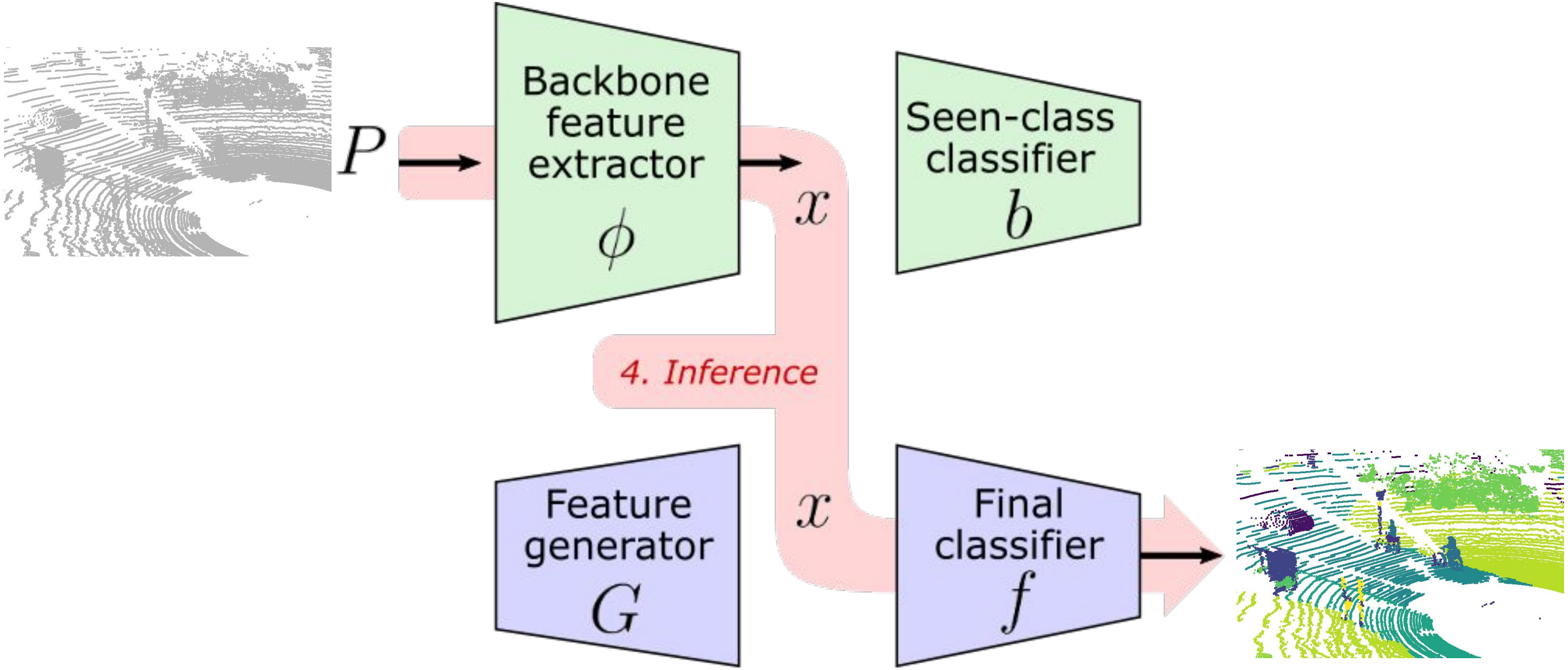
P
Backbone feature extractor
ϕ
x
Seen-class classifier
b
2. Generator training with seen data
z
Class prototypes (auxiliary information)
t_y
Feature generator
G
$\hat{x}$
Final classifier
f

Method



Backbone feature extractor
ϕ
Seen-class classifier
b
z
Class prototypes (auxiliary information)
t_y
Feature generator
G
$\hat{x}$
Final classifier
f
$\hat{y}$
3. Classifier training on generated unseen data (+ seen for GZSL)

Method


P
Backbone feature extractor ϕ
x
Seen-class classifier b
4. Inference
x
Feature generator G
Final classifier f

Experiments - Classification ZSL

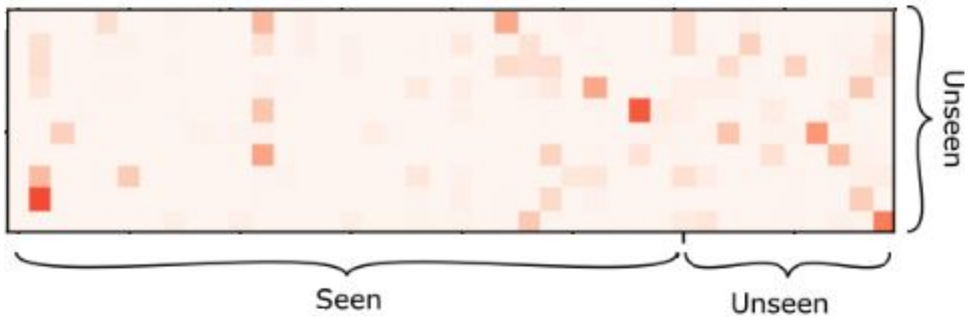
Method	Generative	Full supervision Acc.	ZSL W2V Acc.	ZSL GloVe Acc.	ZSL Glove + W2V Acc.
PointNet		89.2			
f-CLSWGAN[1)]*	✓		20.7	-	-
CADA-VAE[1)]*	✓		23.0	-	-
ZSLPC[2)]			28.0	20.9	20.5
MHPC[3)]			**33.9**	28.7	-
3DGenZ (ours)	✓		28.6	**29.3**	**36.8**

*: adaptation of 2D methods to 3D point clouds, implemented in 1).

1) Cheraghian et al. Transductive zero-shot learning for 3D point cloud classification. WACV, 2020
2) Cheraghian et al. Zero-shot learning of 3d point cloud objects. MVA, 2019.
3) Cheraghian et al. Mitigating the hubness problem for zero-shot learning of 3d objects. 2019.

Bias problem in Generalized Zero-Shot learning (GZSL)

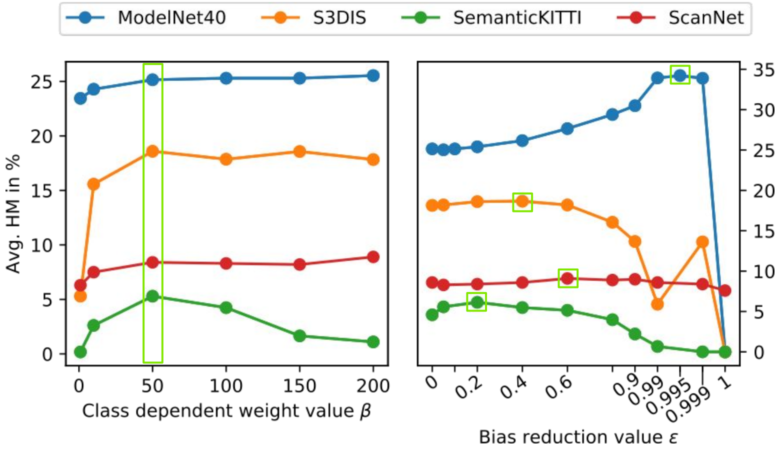
Confusion matrix for unseen classes (GZSL on Modelnet40)



Semantic Segmentation <u>is naturally in GZSL</u>

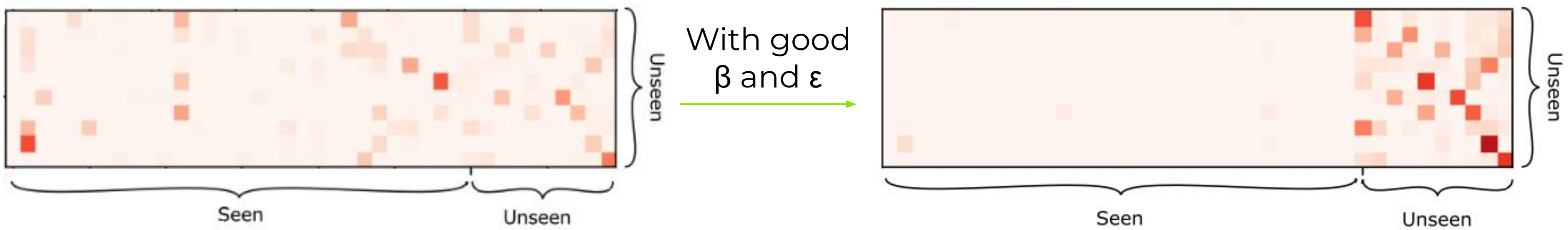
Bias problem and reduction
Additional bias reduction techniques

- **Class-dependant weighting**: Loss for unseen classes is weighted (factor β > 1) in classifier training
- **Calibrated Stacking**: Subtracting a small value ε from the seen-class score (after softmax) at test time
- β and ε are Hyperparameters



Bias problem and reduction
Additional bias reduction techniques

- **Class-dependant weighting**: Loss for unseen classes is weighted (factor $\beta > 1$) in classifier training
- **Calibrated Stacking**[1)]: Subtracting a small value ε from the seen-class score (after softmax) at test time
- β and ε are Hyperparameters



1) Chao et al. An empirical study and analysis of generalized zero-shot learning for object recognition in the wild. ECCV, 2016.

Experiments - Classification GZSL

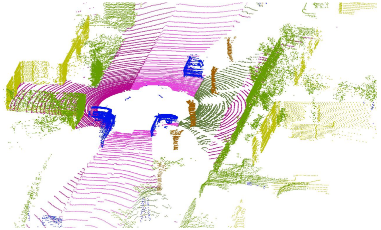
Method	Generative	Full supervision Acc.	Bias reduct.	GZSL W2V Acc. $\mathcal{S}$	GZSL W2V Acc. $\mathcal{U}$	GZSL W2V HM	GZSL GloVe Acc. $\mathcal{S}$	GZSL GloVe Acc. $\mathcal{U}$	GZSL GloVe HM	GZSL GloVe + W2V Acc. $\mathcal{S}$	GZSL GloVe + W2V Acc. $\mathcal{U}$	GZSL GloVe + W2V HM
PointNet		89.2										
f-CLSWGAN[1]*	✓			76.3	3.7	7.0	-	-	-	-	-	-
CADA-VAE[1]*	✓			84.7	1.3	2.6	-	-	-	-	-	-
ZSLPC[2]				40.1	22.5	28.8	49.2	18.2	26.6	-	-	-
MHPC[3]			✓	**53.8**	26.2	35.2	**53.8**	25.7	**34.8**	-	-	-
3DGenZ (ours)	✓		✓	48.8	**29.3**	**36.6**	44.7	**28.4**	34.7	**47.8**	**36.5**	**41.3**

*: adaptation of 2D methods to 3D point clouds, implemented in 1).

1) Cheraghian et al. Transductive zero-shot learning for 3D point cloud classification. WACV, 2020.
2) Cheraghian et al. Zero-shot learning of 3d point cloud objects. MVA, 2019.
3) Cheraghian et al. Mitigating the hubness problem for zero-shot learning of 3d objects. 2019.

Semantic Segmentation - Datasets and baselines

- 3 Datasets
  - Outdoor
    - SemanticKITTI[1)] (4 Unseen, 15 Seen)
  - Indoor
    - S3DIS[2)] (4 Unseen, 9 Seen)
    - ScanNet[3)] (4 Unseen, 16 Seen)



- 2 Baselines
  - Based on DeViSE[4)] (2D) and ZSLPC[5)] (3D classification)
  - Additional bias reduction

1) Behley et al. SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences. ICCV, 2019.
2) Armeni et al. 3D semantic parsing of large-scale indoor spaces. CVPR, 2016.
3) Dai et al. Scannet: Richly-annotated 3d reconstructions of indoor scenes. CVPR, 2017.
4) Frome et al. DeViSE: A deep visual-semantic embedding model. NIPS, 2013.
5) Cheraghian et al. Zero-shot learning of 3d point cloud objects. MVA, 2019.

Experiments Semantic Segmentation - SemanticKITTI

15 seen, 4 unseen classes. Outdoor LiDAR dataset.

	Training set: Backbone	Training set: Classifier	SemanticKITTI mIoU $\mathcal{S}$	SemanticKITTI mIoU $\mathcal{U}$	SemanticKITTI HmIoU
Supervised methods with different levels of supervision					
Full supervision	$\mathcal{S} \cup \mathcal{U}$	$\mathcal{S} \cup \mathcal{U}$	59.4	50.3	54.5
ZSL backbone	$\mathcal{S}$	$\mathcal{S} \cup \mathcal{U}$	52.9	13.2	21.2
ZSL-trivial	$\mathcal{S}$	$\mathcal{S}$	55.8	0.0	0.0
Generalized zero-shot-learning methods					
ZSLPC-Seg[1)]*	$\mathcal{S}$	$\mathcal{U}$	49.1	0.0	0.0
DeViSe-3DSeg[2)]*	$\mathcal{S}$	$\mathcal{U}$	49.7	0.0	0.0
ZSLPC-Seg[1)]	$\mathcal{S}$	$\mathcal{U}$	26.4	10.2	14.7
DeViSe-3DSeg[2)]	$\mathcal{S}$	$\mathcal{U}$	42.9	4.2	7.5
3DGenZ (ours)	$\mathcal{S}$	$\mathcal{S} \cup \hat{\mathcal{U}}$	**41.4**	**10.8**	**17.1**

Upper bounds

*Direct, unrepaired (failing) adaptation.

1) Cheraghian et al. Zero-shot learning of 3d point cloud objects. MVA, 2019.
2) Frome et al. DeViSE: A deep visual-semantic embedding model. NIPS, 2013.

Experiments Semantic Segmentation - SemanticKITTI

15 seen, 4 unseen classes. Outdoor LiDAR dataset.

	Training set Backbone	Training set Classifier	SemanticKITTI mIoU $\mathcal{S}$	SemanticKITTI mIoU $\mathcal{U}$	SemanticKITTI HmIoU
Supervised methods with different levels of supervision					
Full supervision	$\mathcal{S} \cup \mathcal{U}$	$\mathcal{S} \cup \mathcal{U}$	59.4	50.3	54.5
ZSL backbone	$\mathcal{S}$	$\mathcal{S} \cup \mathcal{U}$	52.9	13.2	21.2
ZSL-trivial	$\mathcal{S}$	$\mathcal{S}$	55.8	0.0	0.0
Generalized zero-shot-learning methods					
ZSLPC-Seg[1)]*	$\mathcal{S}$	$\mathcal{U}$	49.1	0.0	0.0
DeViSe-3DSeg[2)]*	$\mathcal{S}$	$\mathcal{U}$	49.7	0.0	0.0
ZSLPC-Seg[1)]	$\mathcal{S}$	$\mathcal{U}$	26.4	10.2	14.7
DeViSe-3DSeg[2)]	$\mathcal{S}$	$\mathcal{U}$	42.9	4.2	7.5
3DGenZ (ours)	$\mathcal{S}$	$\mathcal{S} \cup \hat{\mathcal{U}}$	**41.4**	**10.8**	**17.1**

Baselines

*Direct, unrepaired (failing) adaptation.

1) Cheraghian et al. Zero-shot learning of 3d point cloud objects. MVA, 2019.
2) Frome et al. DeViSE: A deep visual-semantic embedding model. NIPS, 2013.

Experiments Semantic Segmentation - SemanticKITTI

15 seen, 4 unseen classes. Outdoor LiDAR dataset.

	Training set Back-bone	Training set Classi-fier	SemanticKITTI mIoU $\mathcal{S}$	SemanticKITTI mIoU $\mathcal{U}$	SemanticKITTI HmIoU
Supervised methods with different levels of supervision					
Full supervision	$\mathcal{S} \cup \mathcal{U}$	$\mathcal{S} \cup \mathcal{U}$	59.4	50.3	54.5
ZSL backbone	$\mathcal{S}$	$\mathcal{S} \cup \mathcal{U}$	52.9	13.2	21.2
ZSL-trivial	$\mathcal{S}$	$\mathcal{S}$	55.8	0.0	0.0
Generalized zero-shot-learning methods					
ZSLPC-Seg[1]*	$\mathcal{S}$	$\mathcal{U}$	49.1	0.0	0.0
DeViSe-3DSeg[2]*	$\mathcal{S}$	$\mathcal{U}$	49.7	0.0	0.0
ZSLPC-Seg[1]	$\mathcal{S}$	$\mathcal{U}$	26.4	10.2	14.7
DeViSe-3DSeg[2]	$\mathcal{S}$	$\mathcal{U}$	42.9	4.2	7.5
3DGenZ (ours)	$\mathcal{S}$	$\mathcal{S} \cup \hat{\mathcal{U}}$	**41.4**	**10.8**	**17.1**

*Direct, unrepaired (failing) adaptation.

Ours

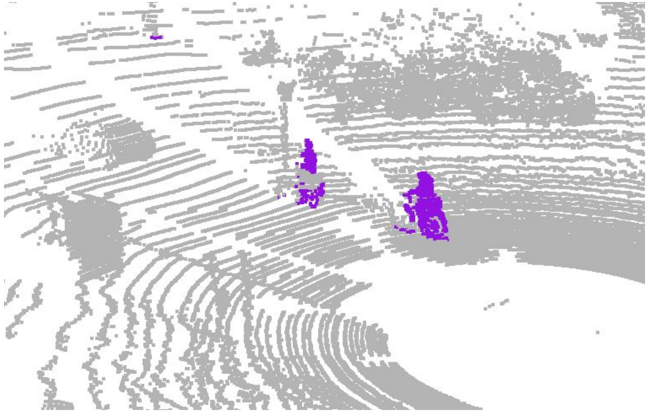
1) Cheraghian et al. Zero-shot learning of 3d point cloud objects. MVA, 2019.
2) Frome et al. DeViSE: A deep visual-semantic embedding model. NIPS, 2013.

Visualisations SemanticKITTI

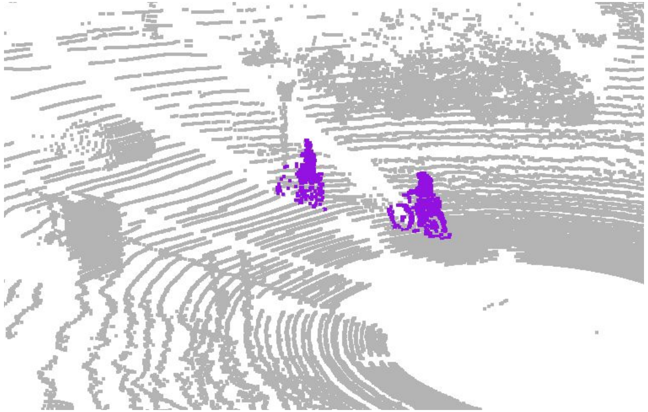
4 unseen, 15 seen classes

Unseen Class
Bicyclist

Predicted



Ground-Truth

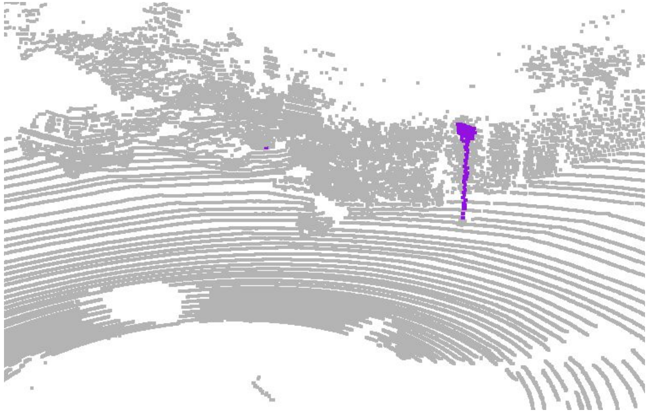


Visualisations SemanticKITTI

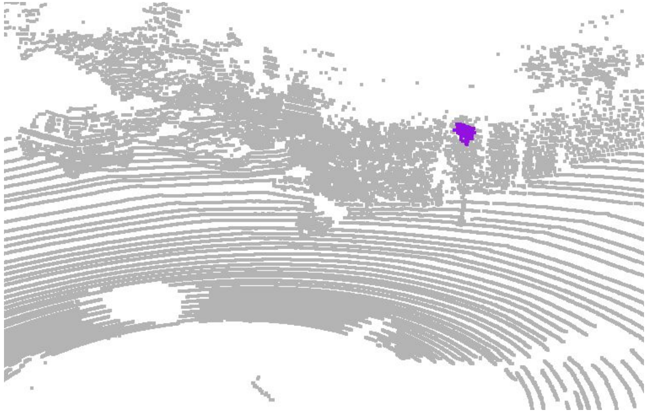
4 unseen, 15 seen classes

Unseen Class
Traffic-Sign

Predicted



Ground-Truth

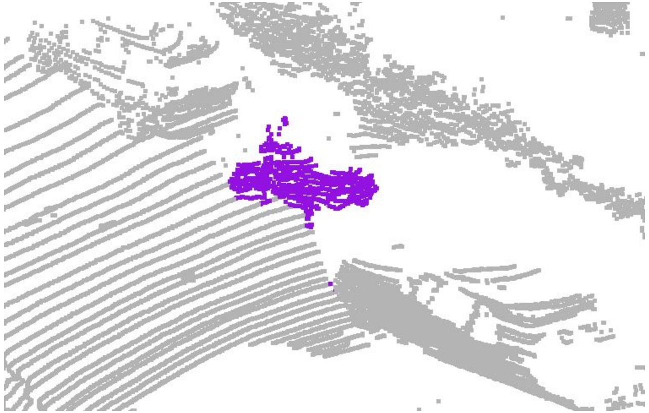


Visualisations SemanticKITTI

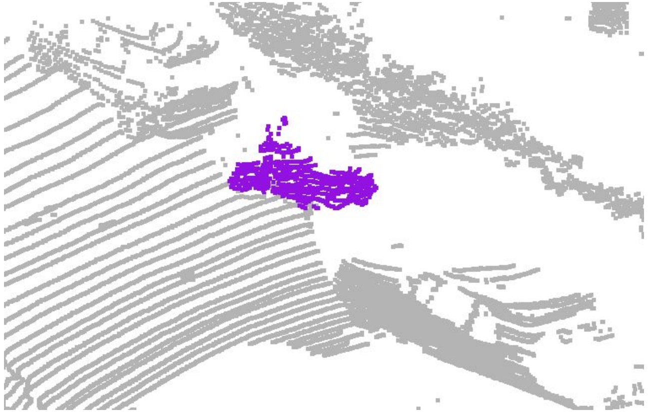
4 unseen, 15 seen classes

Unseen Class
Motorcycle

Predicted



Ground-Truth

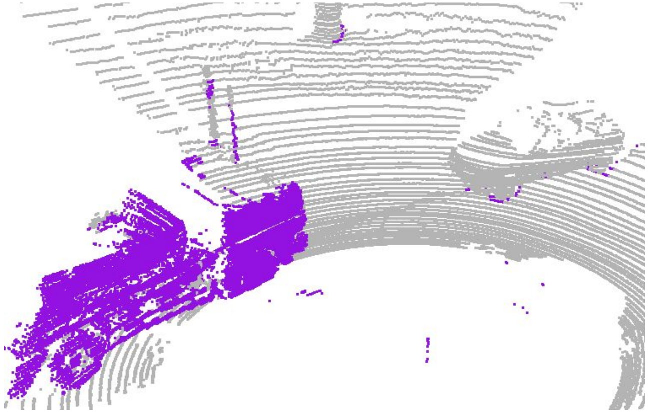


Visualisations SemanticKITTI

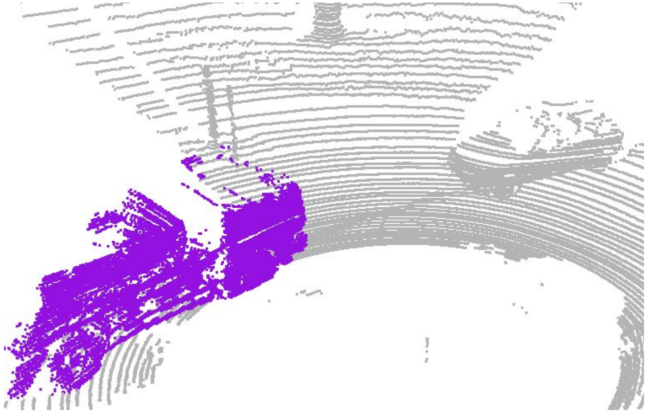
4 unseen, 15 seen classes

Unseen Class
Truck

Predicted



Ground-Truth



Experiments Semantic Segmentation - S3DIS, ScanNet

	Training set		S3DIS			ScanNet		
	Back-bone	Classi-fier	mIoU $\mathcal{S}$	mIoU $\mathcal{U}$	HmIoU	mIoU $\mathcal{S}$	mIoU $\mathcal{U}$	HmIoU
Supervised methods with different levels of supervision								
Full supervision	$\mathcal{S} \cup \mathcal{U}$	$\mathcal{S} \cup \mathcal{U}$	74.0	50.0	59.6	43.3	51.9	47.2
ZSL backbone	$\mathcal{S}$	$\mathcal{S} \cup \mathcal{U}$	60.9	21.5	31.8	41.5	39.2	40.3
ZSL-trivial	$\mathcal{S}$	$\mathcal{S}$	70.2	0.0	0.0	39.2	0.0	0.0
Generalized zero-shot-learning methods								
ZSLPC-Seg[1)]*	$\mathcal{S}$	$\mathcal{U}$	65.5	0.0	0.0	28.2	0.0	0.0
DeViSe-3DSeg[2)]*	$\mathcal{S}$	$\mathcal{U}$	70.2	0.0	0.0	20.0	0.0	0.0
ZSLPC-Seg[1)]	$\mathcal{S}$	$\mathcal{U}$	5.2	1.3	2.1	16.4	4.2	6.7
DeViSe-3DSeg[2)]	$\mathcal{S}$	$\mathcal{U}$	3.6	1.4	2.0	12.8	3.0	4.8
3DGenZ (ours)	$\mathcal{S}$	$\mathcal{S} \cup \hat{\mathcal{U}}$	**53.1**	**7.3**	**12.9**	**32.8**	**7.7**	**12.5**

Baselines

*Direct, unrepaired (failing) adaptation.

1) Cheraghian et al. Zero-shot learning of 3d point cloud objects. MVA, 2019.
2) Frome et al. DeViSE: A deep visual-semantic embedding model. NIPS, 2013.

Experiments Semantic Segmentation - S3DIS, ScanNet

	Training set		S3DIS			ScanNet		
	Back-bone	Classi-fier	mIoU		HmIoU	mIoU		HmIoU
			$\mathcal{S}$	$\mathcal{U}$		$\mathcal{S}$	$\mathcal{U}$	
Supervised methods with different levels of supervision								
Full supervision	$\mathcal{S} \cup \mathcal{U}$	$\mathcal{S} \cup \mathcal{U}$	74.0	50.0	59.6	43.3	51.9	47.2
ZSL backbone	$\mathcal{S}$	$\mathcal{S} \cup \mathcal{U}$	60.9	21.5	31.8	41.5	39.2	40.3
ZSL-trivial	$\mathcal{S}$	$\mathcal{S}$	70.2	0.0	0.0	39.2	0.0	0.0
Generalized zero-shot-learning methods								
ZSLPC-Seg[1)]*	$\mathcal{S}$	$\mathcal{U}$	65.5	0.0	0.0	28.2	0.0	0.0
DeViSe-3DSeg[2)]*	$\mathcal{S}$	$\mathcal{U}$	70.2	0.0	0.0	20.0	0.0	0.0
ZSLPC-Seg[1)]	$\mathcal{S}$	$\mathcal{U}$	5.2	1.3	2.1	16.4	4.2	6.7
DeViSe-3DSeg[2)]	$\mathcal{S}$	$\mathcal{U}$	3.6	1.4	2.0	12.8	3.0	4.8
3DGenZ (ours)	$\mathcal{S}$	$\mathcal{S} \cup \hat{\mathcal{U}}$	**53.1**	**7.3**	**12.9**	**32.8**	**7.7**	**12.5**

Ours

*Direct, unrepaired (failing) adaptation.

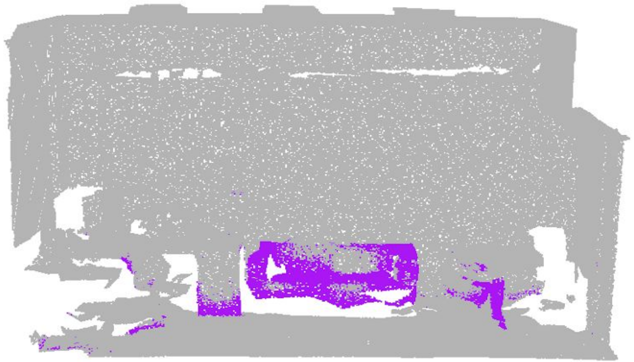
1) Cheraghian et al. Zero-shot learning of 3d point cloud objects. MVA, 2019.
2) Frome et al. DeViSE: A deep visual-semantic embedding model. NIPS, 2013.

Visualisations S3DIS

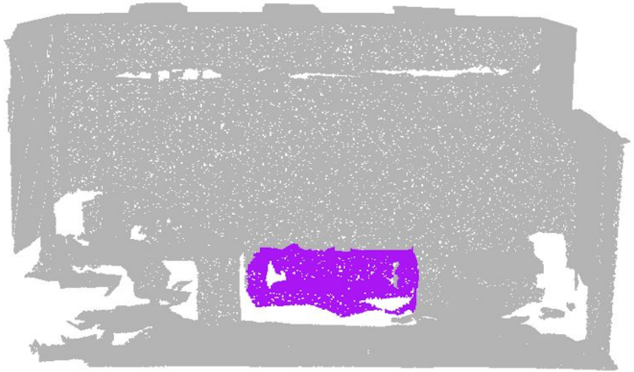
4 unseen, 9 seen classes

Unseen Class
Sofa

Predicted



Ground-Truth

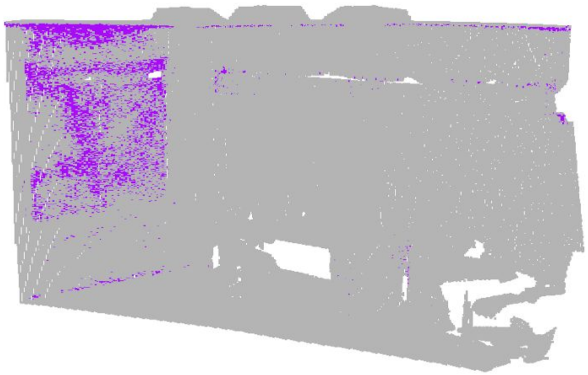


Visualisations S3DIS

4 unseen, 9 seen classes

Unseen Class
Window

Predicted



Ground-Truth

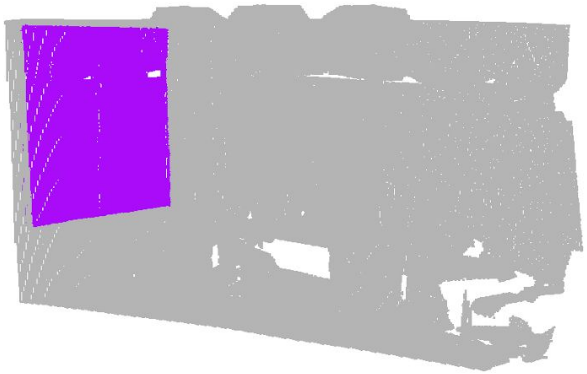


Image based 3D Zero-shot

- Image-based class prototypes

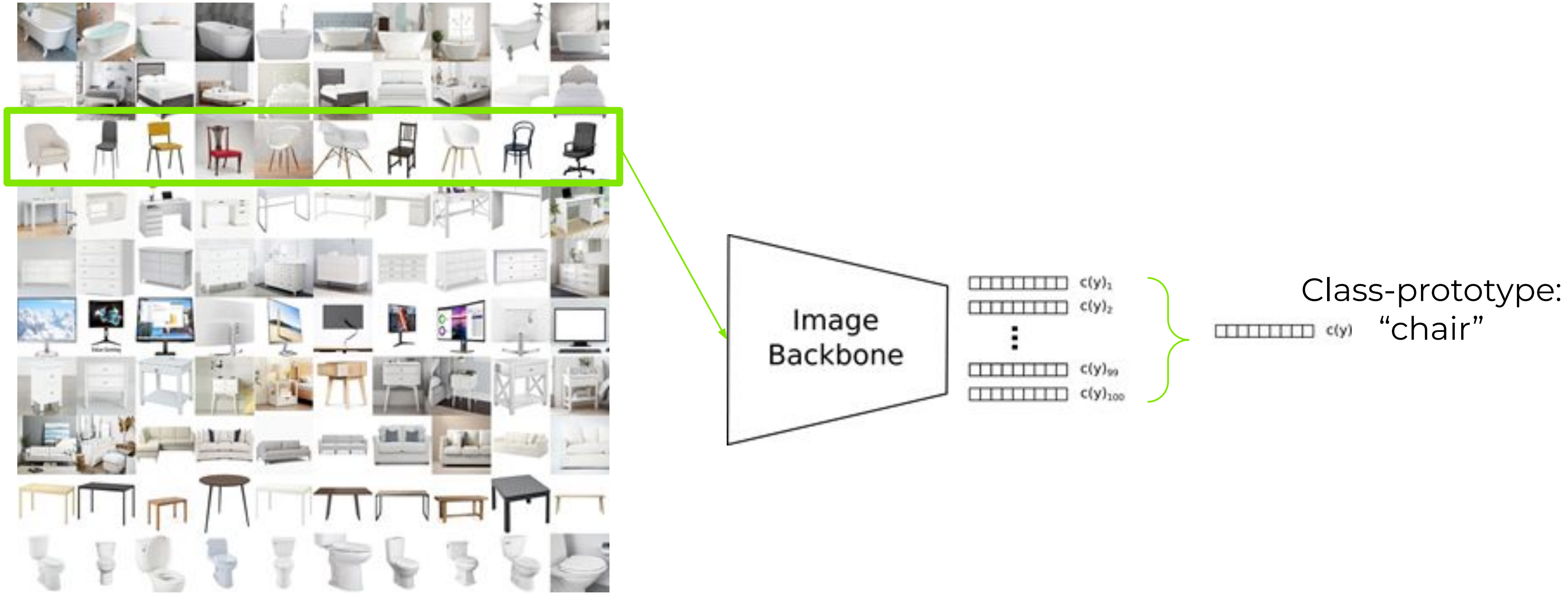
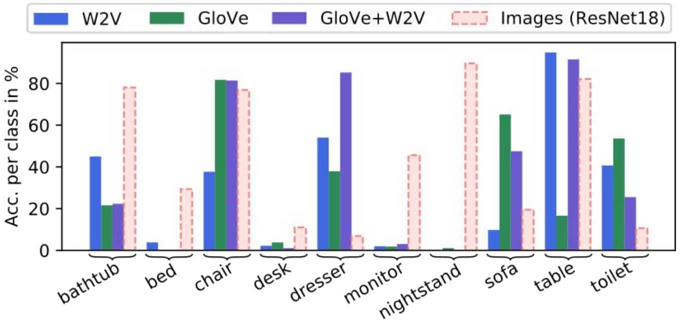


Image based 3D zero-shot - Results

Representation	*Classif.* ModelNet40 ZSL	*Classif.* ModelNet40 GZSL	*Segmentation* ScanNet HmIoU	*Segmentation* KITTI HmIoU
W2V+GloVe (self-sup.)	36.8	**41.3**	12.5	**17.1**
ResNet-18 (sup.)	**43.6**	40.0	13.9	3.6
ResNet-50 (self-sup.)	37.0	36.5	**15.5**	5.3

Only image based



Conclusion and outview

Conclusion

1. Reaching state of the art in **Classification** with additional bias reduction
2. GZSL for **Semantic Segmentation** on 3D PCs for 3 datasets:
   a. Improving over natural baselines
   b. Creation of a benchmark
3. **Image-based class prototypes** can outperform text-based ones

Outview

- Multi-Modal prototypes
- Phrasal (multi-word) embeddings to discover complex corner cases

SMART TECHNOLOGY
FOR SMARTER MOBILITY