Sparsely Annotated Semantic Segmentation with Adaptive Gaussian Mixtures

Linshan Wu[1], Zhun Zhong[2], Leyuan Fang[1], Xingxin He[1], Qiang Liu[1], Jiayi Ma[3], Hao Chen[4]
Hunan University[1]   University of Trento[2]   Wuhan University[3]   HKUST[4]

Sparsely Annotated Semantic Segmentation with Adaptive Gaussian Mixtures

Linshan Wu[1], Zhun Zhong[2], Leyuan Fang[1*], Xingxin He[1], Qiang Liu[1], Jiayi Ma[3], Hao Chen[4]
Hunan University[1] University of Trento[2] Wuhan University[3] HKUST[4]

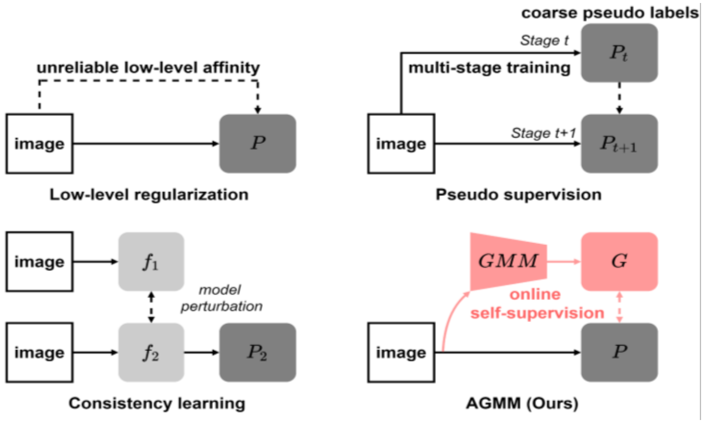
Sparsely Annotated Semantic Segmentation

➢ **What?**

Using low-cost sparse labels (points and scribbles) instead of expensive pixel-level labels for supervision

➢ **Problems?**

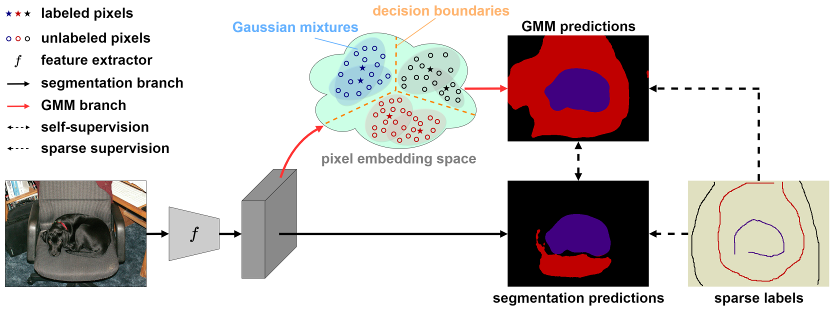
- Existing methods are mainly based on unreliable low-level information and coarse pseudo labels
- More reliable supervision need to be introduced



➢ **Motivation**

Modeling the distributions between the labeled and unlabeled pixels to generate reliable pseudo labels online for dynamic self-supervision

Adaptive Gaussian Mixture Model

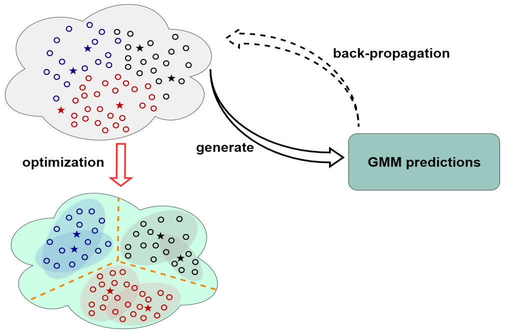


➢ Reliable labeled pixels as the centers of class-wise Gaussian Mixtures

GMM formulations and Self-supervision loss:

$$G = \sum_{i}^{K} g_i(x_i, \mu_i, \sigma_i) = \sum_{i}^{K} e^{-\frac{d^2}{2\sigma_i^2}}$$

$$L_{self} = -\frac{1}{|x|}\sum[G * log(p) + (1-G) * \log(1-p)]$$



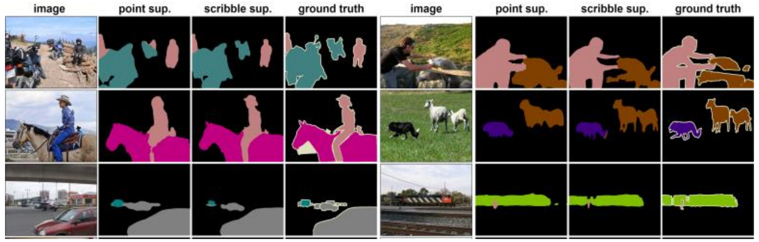
➢ Optimizing AGMM through back-propagation

➢ Learning discriminative decision boundaries

Experiments

Point sup.	Pub.	Back.	Mul.	CRF	Val
KerCut	ECCV18	R101	√	√	57.0
SEAM	CVPR20	R101	√	√	66.3
A2GNN	PAMI21	R101	√	√	66.8
Seminar	ICCV21	R101	√	√	72.5
SPML	ICLR	R101	-	√	73.2
DBFNet	TIP22	R101	-	-	66.8
TEL	CVPR22	R101	-	-	64.9
AGMM	CVPR23	R101	-	-	**69.6**

Scrib. sup.	Pub.	Back.	Mul.	CRF	Val
URSS	ICCV21	R101	√	√	76.1
PSI	ICCV21	R101	-	-	74.9
Seminar	ICCV21	R101	√	-	76.2
A2GNN	PAMI21	R101	√	√	74.3
DBFNet	TIP22	R101	-	-	72.5
TEL	CVPR22	R101	-	-	75.8
AGMM	CVPR23	R101	-	-	**76.4**



Summary

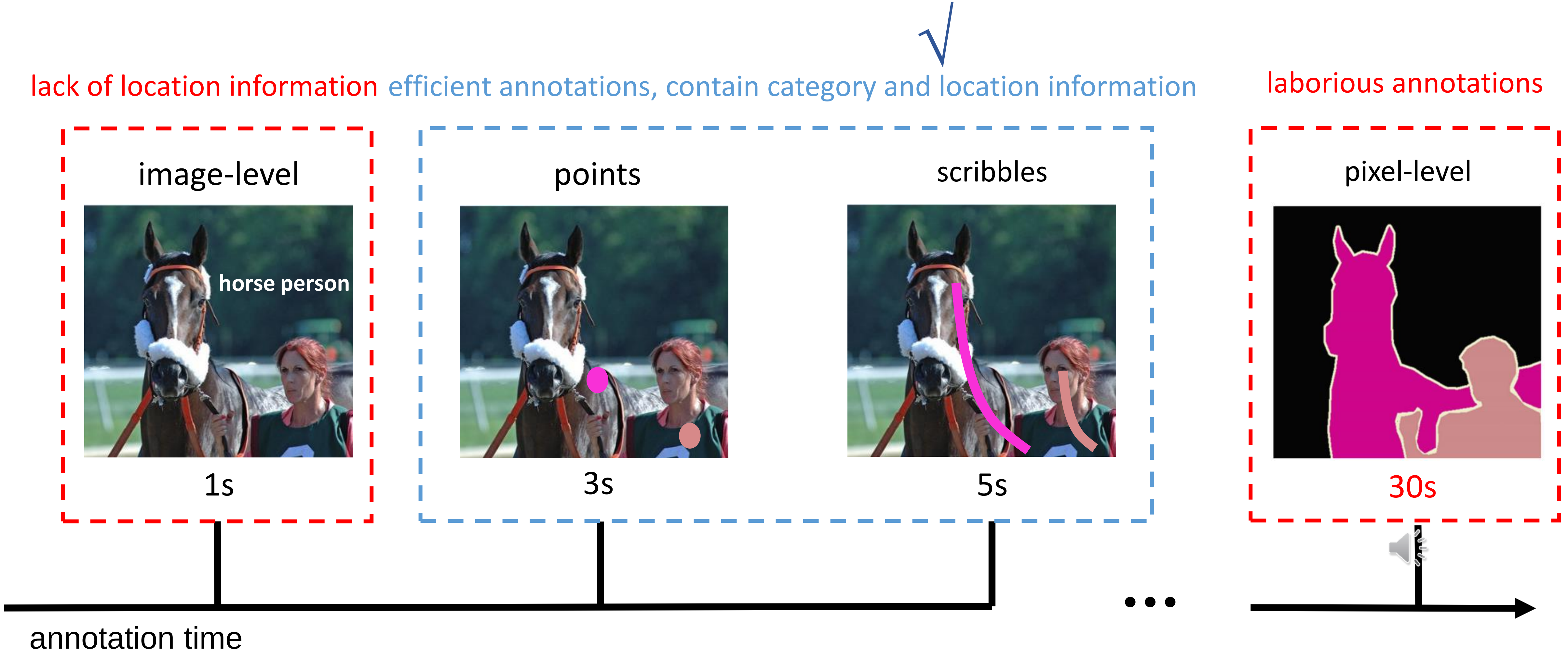
➢ We achieve state-of-the-art performances in SASS.

➢ **Code available:** https://github.com/Luffy03/AGMM-SASS

Background

- **Sparsely annotated semantic segmentation (SASS)**

  Use sparse labels (points or scribbles) for supervision



Solutions for SASS

(a) Low-level regularization

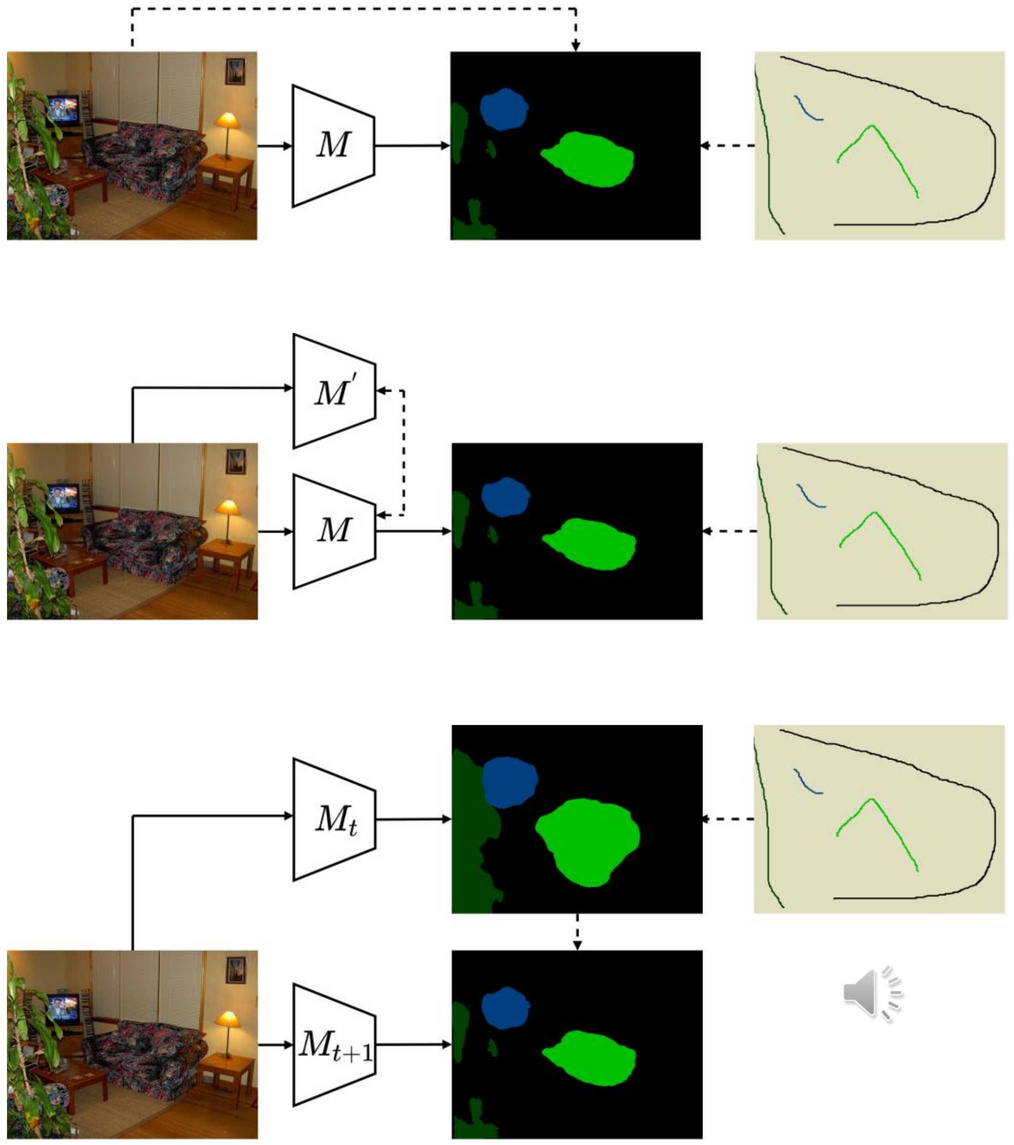
- Use the low-level affinity for supervision
- Ignore the large gap between low-level visuals and high-level semantics

(b) Consistency learning

- Learn consistent features from different views
- Fail to supervise the final predictions in the category-level

(b) Pseudo supervision

- Generate pseudo labels for supervision
- Time-consuming multi-stages training
- Pseudo labels generation are not reliable

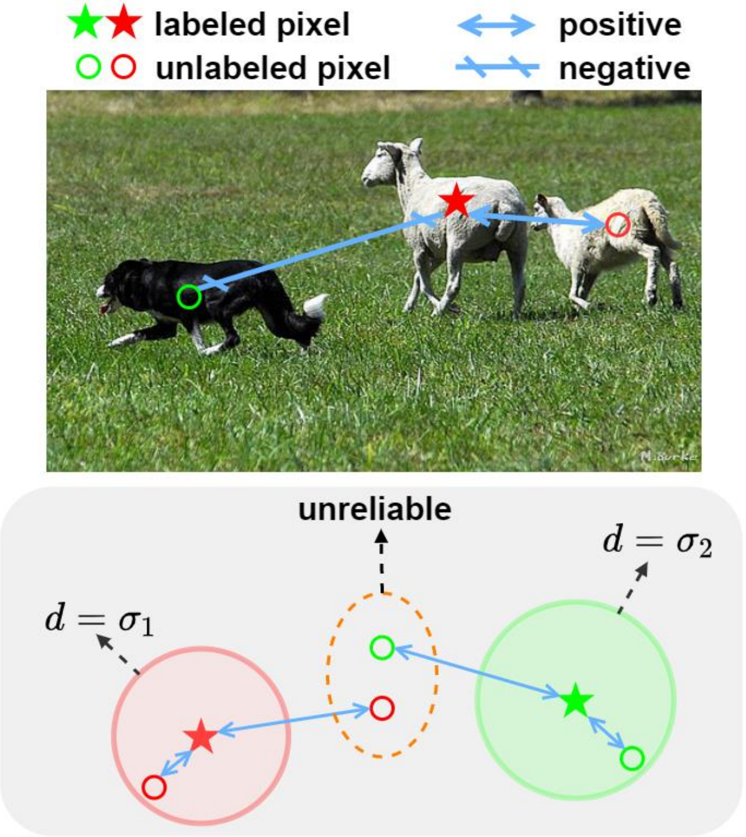


Motivation

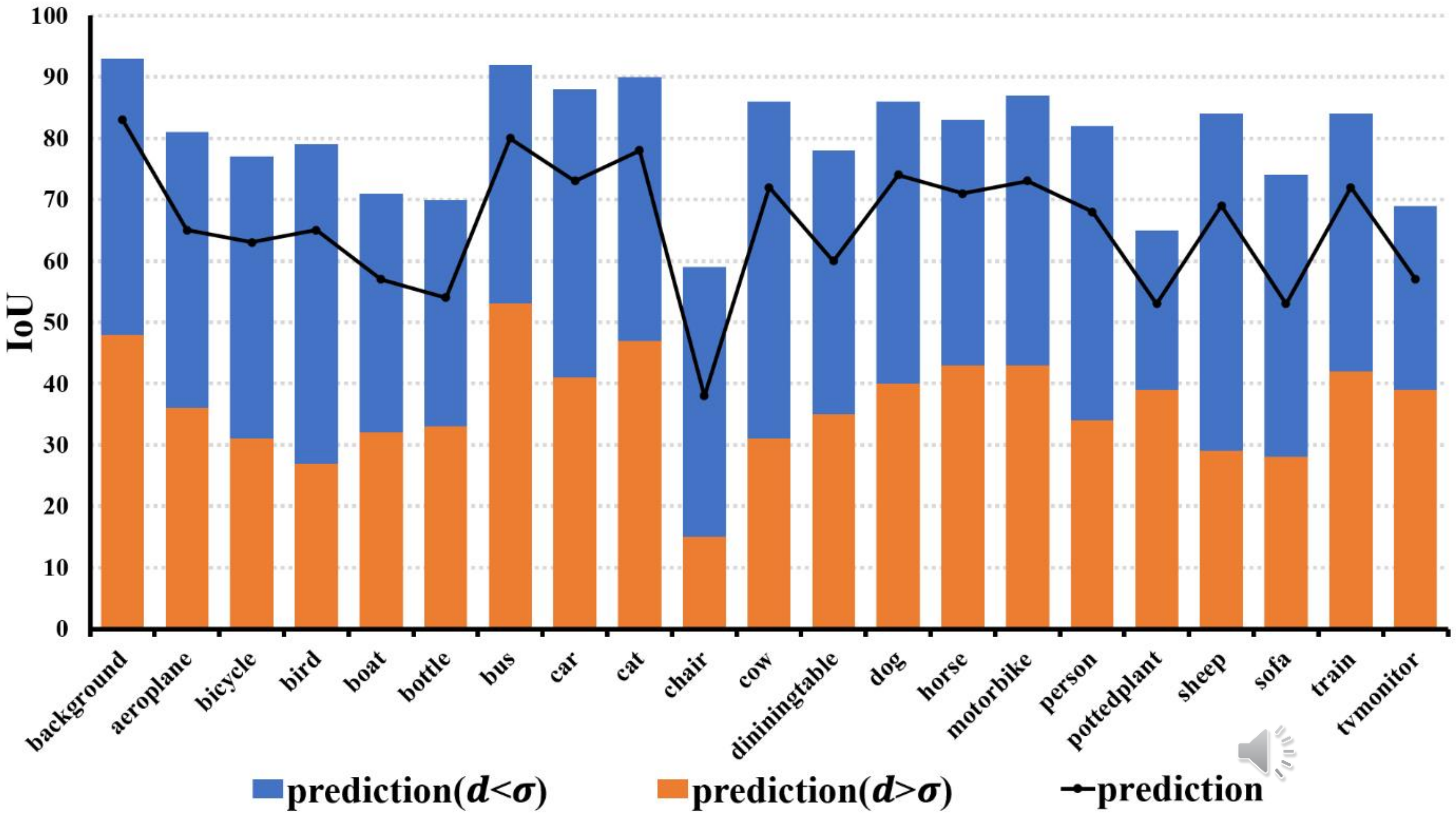
Introduce more reliable information for supervision

(1) We observe that the similarity between labeled and unlabeled pixels is highly associated with the predictions of unlabeled pixels

(2) We emphasize that reliable information of labeled pixels can be explored to supervise the unlabeled pixels



(a) Relation between labeled and unlabeled pixels

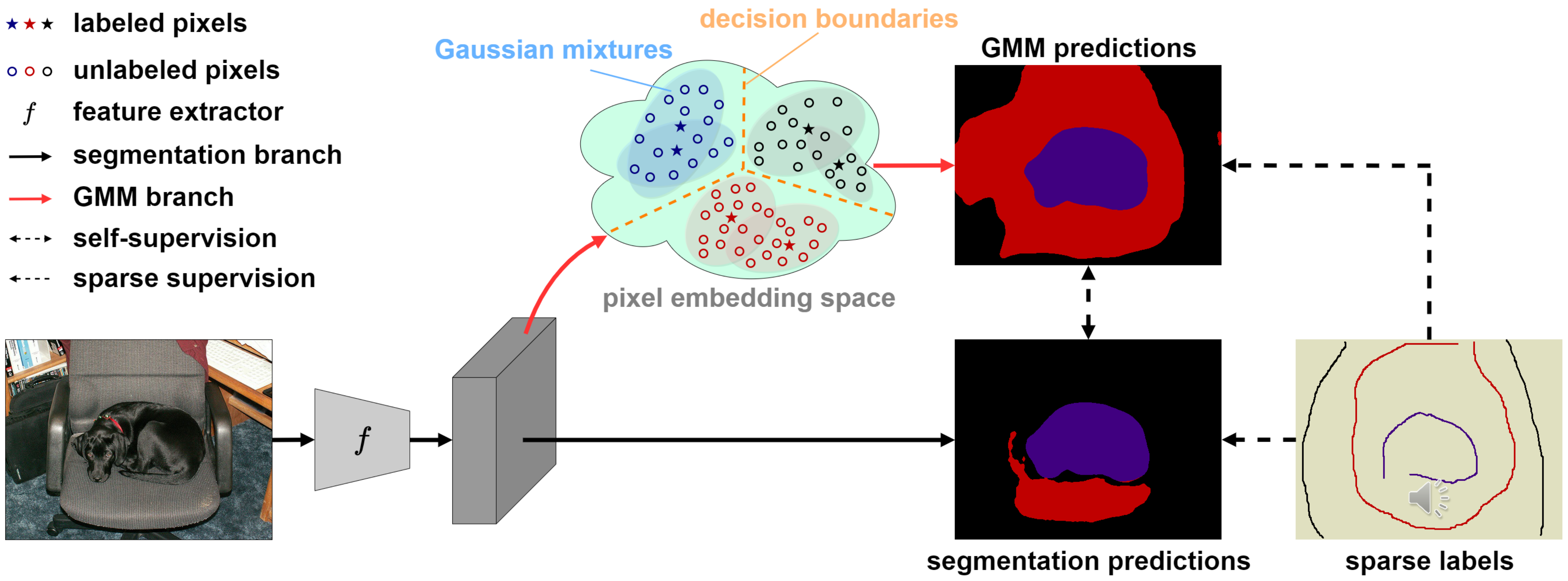


(b) Category-wise performance

Adaptive Gaussian Mixtures Model (AGMM)

(1) We build a GMM branch to generate GMM predictions for supervision

(2) GMM is formulated by modeling the distributions among labeled and unlabeled pixels



GMM formulation

Assign the labeled pixels as the centers μ_i of i_{th} category-wise Gaussian Mixtures:

$$\mu_i = \frac{1}{|x_{li}|} \sum_{\forall x \in x_{ls}} f(x)$$

Calculate the variance σ_i with predictions P_i and means μ_i:

$$\sigma_i = \sqrt{\frac{1}{|P_i|} \sum_{\forall x \in x_u} P_i d^2}$$

Where d is the distance between the features of pixels x and means μ_i:

$$d = |f(x) - \mu|$$

Then, we can bulid GMM G for K classes as follows:

$$G = \sum_{i}^{K} g(x, \mu_i, \sigma_i) = \sum_{i}^{K} e^{-\frac{d^2}{2\sigma_i^2}}$$

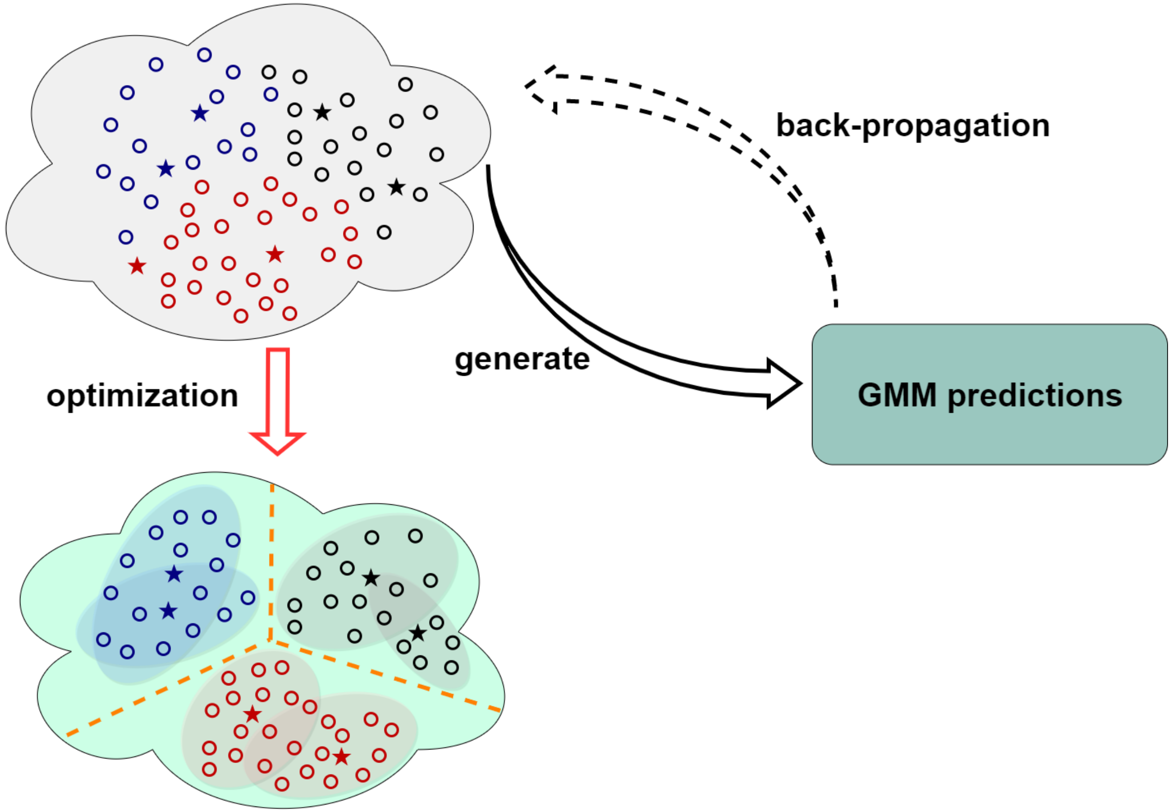
GMM supervision

Online self-supervision between GMM and segmentation predictions, G and P:

$$L_{self} = -\frac{1}{|x|}\sum[G * log(p) + (1-G) * \log(1-p)]$$

(1) Optimizing AGMM through back-propagation

(2) Learning discriminative decision boundaries



Strong performance

AGMM outperform existing state-of-the-art methods by a large margin

Point-supervised SASS on PASCAL VOC 2012 dataset

Method	Pub.	Backbone	Extra Data	Multi-stage	CRF	Val
What point	ECCV16	VGG16	-	-	-	43.4
KernerCut	ECCV18	R101	-	√	√	57.0
SEAM	CVPR20	R101	-	√	√	66.3
A2GNN	PAMI21	R101	-	√	√	66.8
Seminar	ICCV21	R101	-	√	√	72.5
SPML	ICLR	R101	√	-	√	73.2
DBFNet	TIP22	R101	-	-	-	66.8
TEL	CVPR22	R101	-	-	-	64.9
AGMM	CVPR23	R101	-	-	-	**69.6**

Strong performance

AGMM outperform existing state-of-the-art methods by a large margin

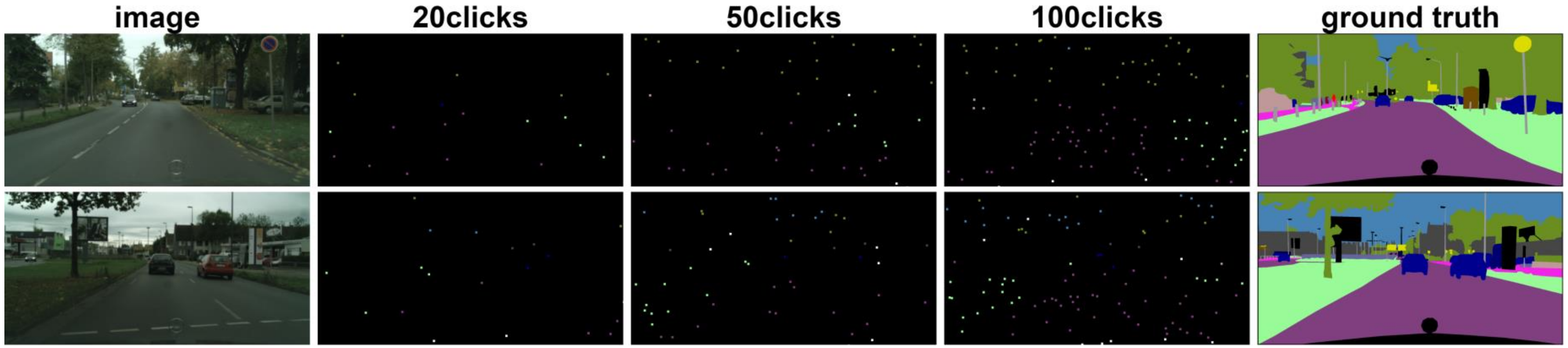
Scribble-supervised SASS on PASCAL VOC 2012 dataset

Method	Pub.	Backbone	Extra Data	Multi-stage	CRF	Val
ScribbleSup	CVPR16	VGG16	-	√	√	63.1
RAWKS	CVPR17	VGG16	-	√	√	73.5
GraphNet	ACMM18	R101	-	√	√	74.5
NormCut	CVPR18	R101	-	√	√	75.0
GridCRF	CVPR19	R101	-	-	-	72.8
SEAM	CVPR20	R101	-	√	√	75.0
BPG	IJCAI19	R101	√	-	-	76.0
SPML	ICLR21	R101	√	-	√	76.1
URSS	ICCV21	R101	-	√	√	76.1
PSI	ICCV21	R101	-	-	-	74.9
Seminar	ICCV21	R101	-	√	-	76.2
A2GNN	PAMI21	R101	-	√	√	74.3
DBFNet	TIP22	R101	-	-	-	72.5
TEL	CVPR22	R101	-	-	-	75.8
AGMM	CVPR23	R101	-	-	-	**76.4**

Strong performance

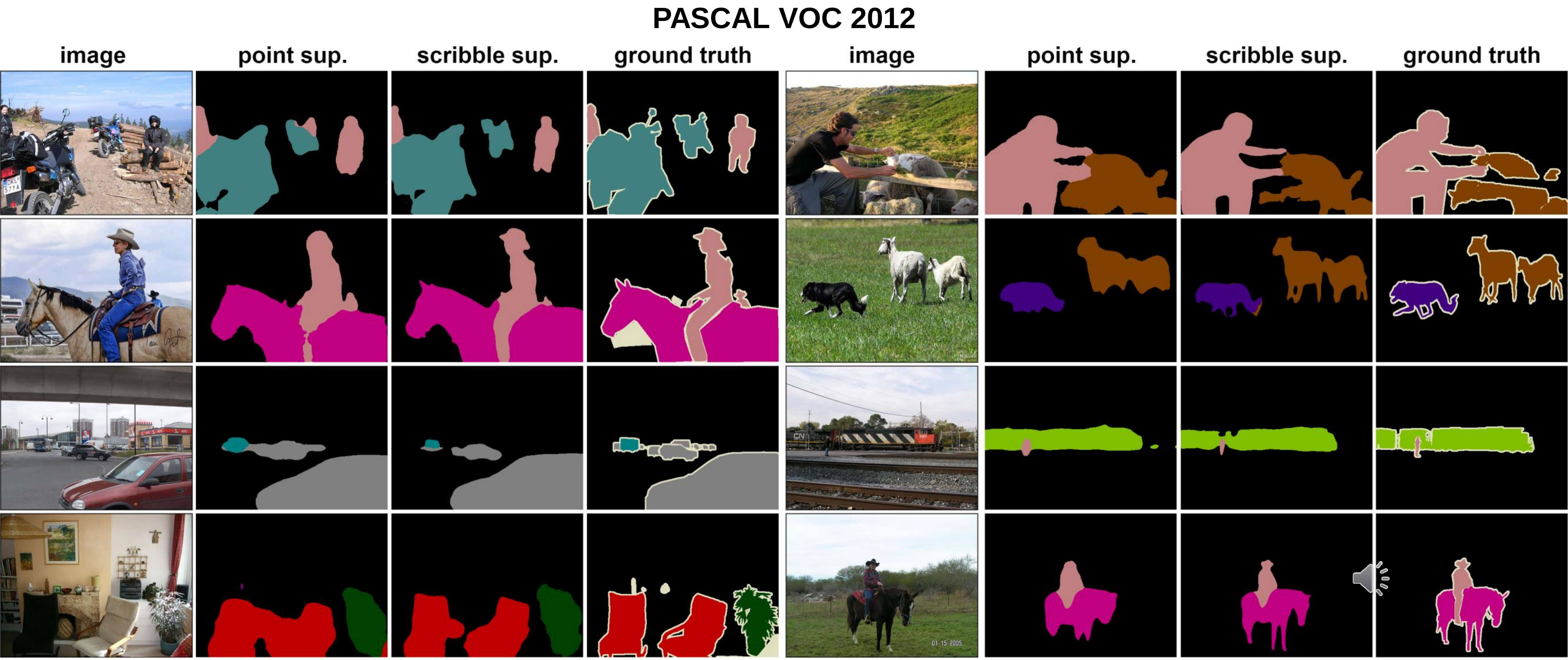
AGMM outperform existing state-of-the-art methods by a large margin

Point-supervised SASS on Cityscapes dataset



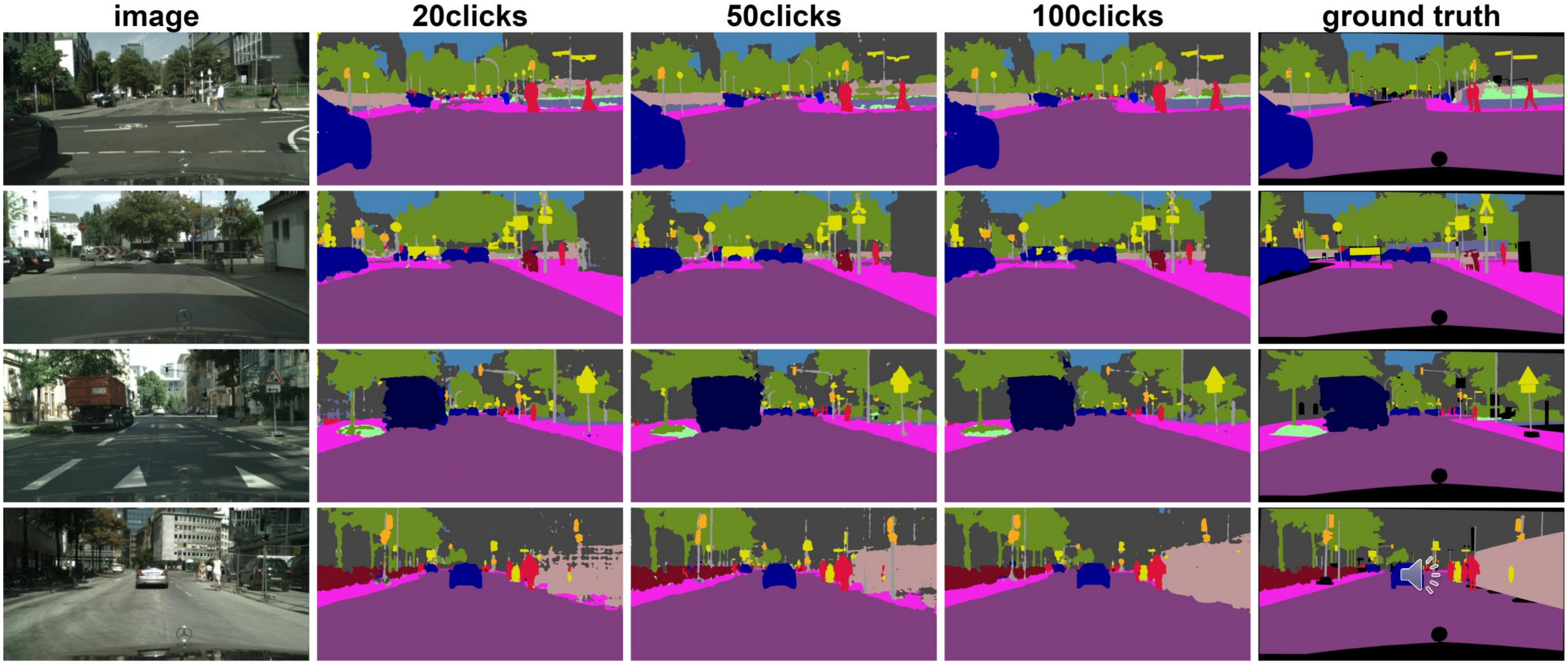
Method	Cityscapes			
	20 clicks	50 clicks	100 clicks	full
Baseline	53.5	60.3	64.2	78.6
DenseCRF	54.2	61.6	65.5	-
Seminar	57.1	63.0	66.1	-
TEL	56.3	62.8	67.6	-
AGMM	**76.1**	**68.3**	**71.6**	-

Qualititive Visualization Results



Qualititive Visualization Results

Cityscapes



Conclusion & Future Work

- Conclusion
  - We proposed a simple yet effective framework AGMM for SASS. Specifically, we assigned the labeled pixels as the centroids of category-wise Gaussian mixtures, enabling us to formulate a GMM to model the similarity between labeled and unlabeled pixels. Then, we can leverage the reliable information from labeled pixels to generate GMM predictions for dynamic online selfsupervision. Extensive experiments demonstrate our method achieves state-of-the-art SASS performance..

- Future Work
  - Explore the AGMM for unified Weakly-Supervised Semantic Segmentation (WSSS), e.g., image-level labels, bounding-box labels.

Thank you for listening!