CVPR 2025

Toward Real-world BEV Perception: Depth Uncertainty Estimation via Gaussian Splatting

Shu-Wei Lu[1] Yi-Hsuan Tsai[2] Yi-Ting Chen[1]*
[1]National Yang Ming Chiao Tung University [2]Atmanity Inc.

- Problem/Objective
  - BEV segmentation + GS

- Contribution/Key Idea
  - Novel depth uncertainty modeling approach
    - Leverages depth ambiguity
  - Computationally efficient by GS
  - SOTA (among 2D unprojection approach)
    - Significant memory reduce

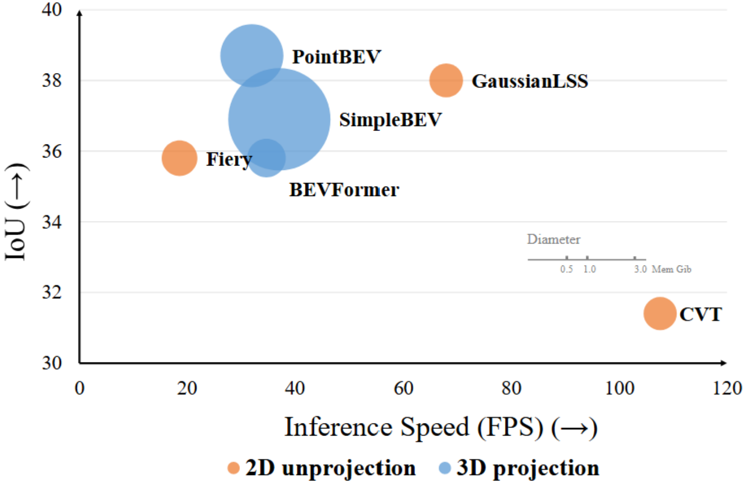


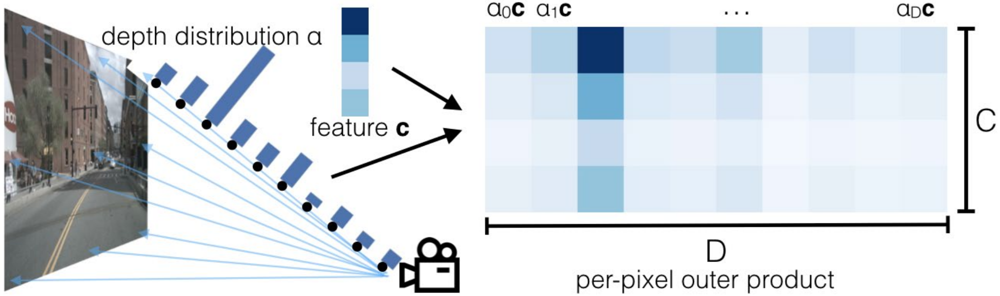
Figure 1. **Comparisons of 2D unprojection-based and 3D projection-based methods on vehicle BEV segmentation.** GaussianLSS achieves state-of-the-art performance among 2D unprojection baselines. In addition, it also demonstrates competitive performance compared to 3D projection-based methods, while offering significant advantages in memory efficiency and inference speed.

김범준

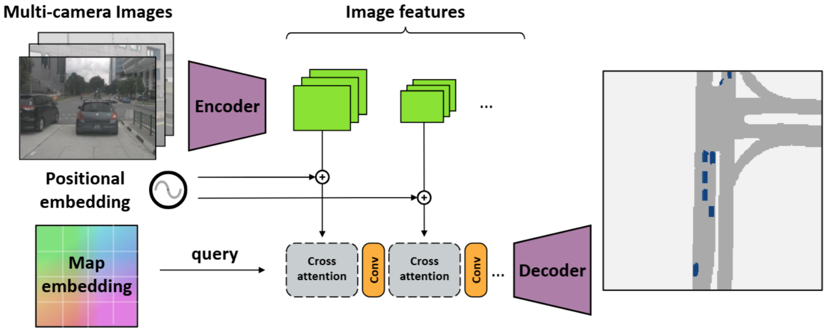
Toward Real-world BEV Perception: Depth Uncertainty Estimation via Gaussian Splatting

CVPR 2025

Introduction - BEV


depth distribution α
feature **c**
$\alpha_0\mathbf{c}$ $\alpha_1\mathbf{c}$ ... $\alpha_D\mathbf{c}$
C
D
per-pixel outer product

Depth-based method



Attention-based method

- 3D projection
  - 3D 공간을 미리 정해놓고 projection
    - ex) BEVFormer, SimpleBEV
    - depth prediction의 complexity를 간과
    - Computational complexity

- Implicit 2D unprojection
  - image → BEV **without predicting depth**
    - ex) CVT, Petrv2
    - BEV grid가 커짐에 따라 computation ↑

- Explicit 2D unprojection
  - Relies heavily on accurate **depth estimation**
    - ex) LSS, BEVFusion
    - depth supervision 등을 추가했지만 효과x
    - **명시적인 깊이 불확실성 표현이 부족**
    - **→ 복잡한 상황에서 depth ambiguity를 제대로 처리 x**

Introduction - Uncertainty modeling

- Variance of Predicted Distributions
  - 각 예측 값이 확률 분포로 나오는 경우
  - 불확실성을 분산을 통해 예측
    - 분포가 넓으면(분산 큼) 불확실성이 크다고 해석

→ 해당 논문에서는 distribution의 variance를 이용

→ by depth 분포(softmax over depth bins)의 분산(variance)

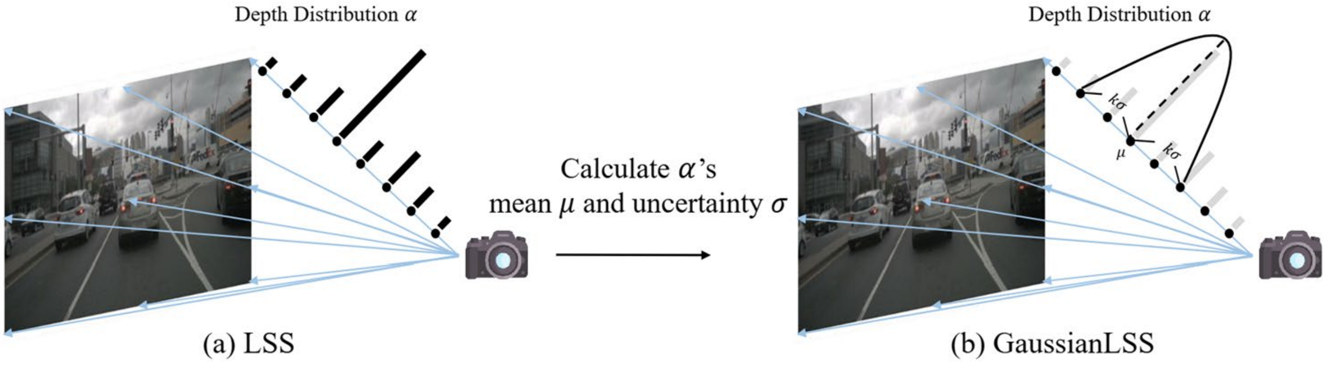
- MLP-Based
  - output: single uncertainty score
    - 불확실성 점수(혹은 분포: 평균+분산)를 직접 출력

- Bayesian Network
  - 불확실성(예: posterior, 혹은 variance)을 이론적으로 유도
  - 파라미터에 prior 분포를 두고, 예측 결과의 variance를 뽑음

CVPR 2025

Toward Real-world BEV Perception: Depth Uncertainty Estimation via Gaussian Splatting

● Method - Main contribution


Depth Distribution α
Calculate α's
mean μ and uncertainty σ
Depth Distribution α
$k\sigma$
μ
$k\sigma$

(a) LSS

(b) GaussianLSS

- LSS 방식의 한계 (depth estimation 방법)
  - 분포의 모양이 sharp, ambiguous하면 depth 오류가 BEV에 직접 전파됨
  - "depth ambiguity problem" (깊이 추정 자체의 불확실성이 BEV representation을 ↓
- **GaussianLSS**
  - 분포의 평균(μ, mean)과 분산(σ, uncertainty)을 계산
    - [μ−kσ, μ+kσ]라는, uncertainty-aware(불확실성 반영) "범위"를 정의
    - k: error tolerance coefficient, spread의 폭을 조절
  - *“확률적으로 이 정도 범위에 있음”* 을 smooth, continuous하게 정의

김범준

- **Method - Framework**

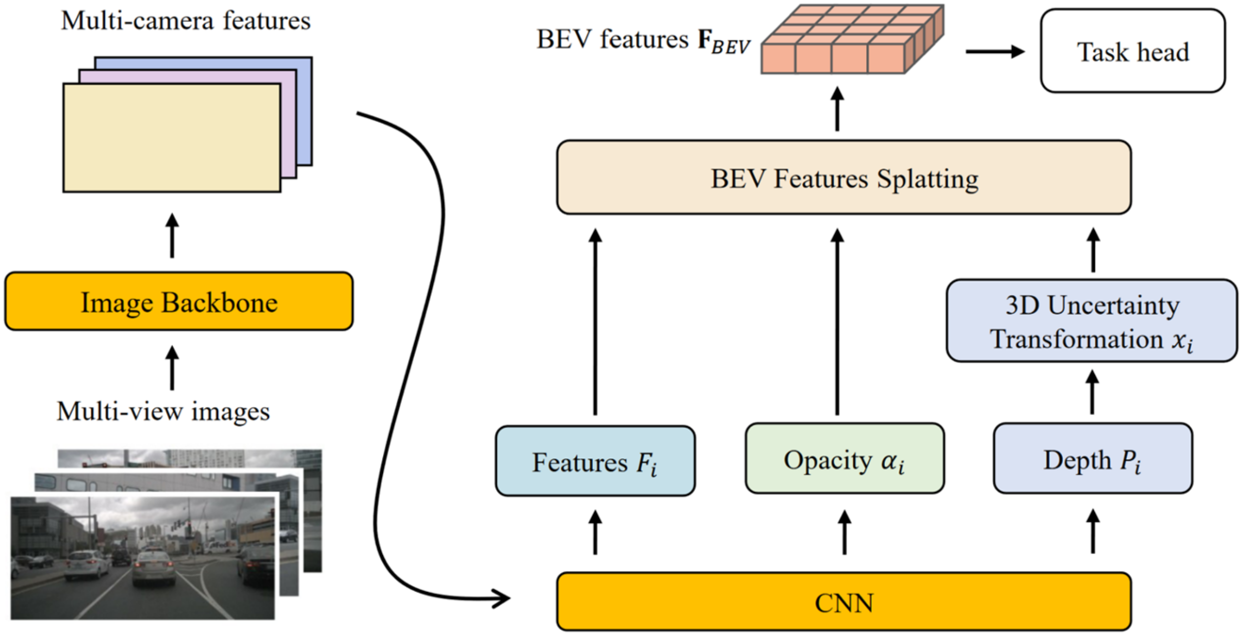


Figure 3. **Overview of GaussianLSS.** Multi-view images are first processed by a backbone network to extract features. They are then input to a simple CNN layer to obtain splat features F_i, opacity α_i, and depth distribution P_i. The predicted depth distribution undergoes an uncertainty transformation to produce a 3D uncertainty x_i. Next, BEV features are obtained through a splatting process, integrating features across views. The resulting BEV features $\mathbf{F}_{\text{BEV}}$, enriched with uncertainty awareness, are used as input to the task-specific head for prediction.

Toward Real-world BEV Perception: Depth Uncertainty Estimation via Gaussian Splatting

CVPR 2025

- ## Method - GaussianLSS

(LSS 구조)

$$D = \left\{ d_i = d_{\min} + i \cdot \frac{d_{\max} - d_{\min}}{B} \right\}_{i=0}^{B-1}$$

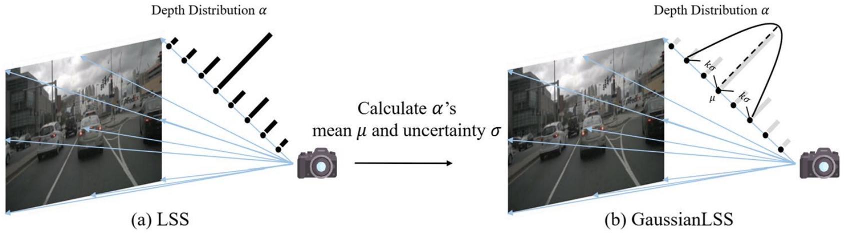
$$\downarrow$$

$$\mathcal{C} \in \mathbb{R}^{H \times W \times B \times 3}$$

$$\downarrow$$

$$\mathbf{c}_d = \alpha_d \mathbf{c}$$

depth distribution coefficient α_d



Depth Distribution α
Calculate α's
mean μ and uncertainty σ
Depth Distribution α
$k\sigma$
μ
$k\sigma$

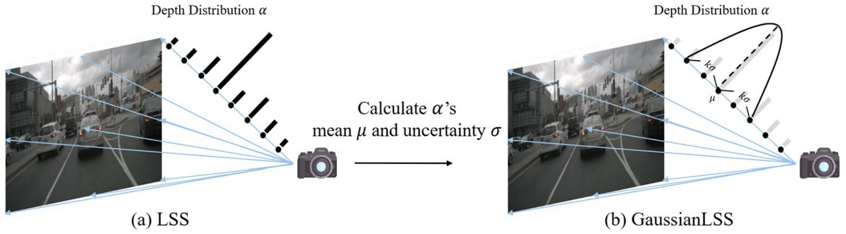
(a) LSS

(b) GaussianLSS

- LSS 방식의 한계 (depth estimation 방법)
  - Sparse BEV projection
    - discrete하게 예측
  - Unstable depth distribution
    - softmax는 인접 bin 사이에서도 확률 크게 차이 가능
    - → 유사한 깊이도 BEV feature가 일관성 없는 분배 가능
- GaussianLSS
  - Smooth & reliable 한 BEV feature aggregation 가능

김범준

Method - GaussianLSS


Depth Distribution α
Depth Distribution α
Calculate α's
mean μ and uncertainty σ
$k\sigma$
μ
$k\sigma$

(a) LSS

(b) GaussianLSS

- 픽셀마다 가지는 depth distribution
  - 평균, 분산 계산

for each pixel p

$$\mu = \sum_{i=0}^{B-1} P_i(p) d_i$$

$$\sigma^2 = \sum_{i=0}^{B-1} P_i(p)(d_i - \mu)^2$$

$$\downarrow$$

$$\hat{\mathbf{D}} = [\mu - k\sigma, \mu + k\sigma]$$

기존 LSS 확률 제일 높은 depth로 추정 / GaussianLSS는 [μ−kσ ~ μ+kσ] 범위 어딘가에 존재한다고 가정

Toward Real-world BEV Perception: Depth Uncertainty Estimation via Gaussian Splatting

CVPR 2025

- ## Method - 3D Uncertainty Transformation

$$x_i = T(\hat{\mathbf{D}}, I, E) = (\mu_{3d}, \Sigma)$$

: **이 픽셀의 3D 위치와 그 불확실성(공간 spread)"을 계산**하는 함수 T

- 이미지 좌표계(2D + depth) → 3D world(Ego-vehicle) 좌표계

$$p = (u, v, d)$$

↓

$$\mu_{3d} = \sum_{i=0}^{B-1} P_i(p)\, p_i^{3d},$$

모든 depth 후보에 대해 확률적 평균(가중합) → 3D Gaussian의 평균(center)

: *"이 픽셀 feature가 가장 있을 법한 3D center"*

$$\Sigma = \sum_{i=0}^{B-1} P_i(p)\, (p_i^{3d} - \mu_{3d})(p_i^{3d} - \mu_{3d})^T,$$

Σ 3D Gaussian의 공분산 행렬

: *“가우시안이 퍼진 정도”*

$$(\mathbf{x} - \mu_{3d})^T \Sigma^{-1} (\mathbf{x} - \mu_{3d}) \leq k^2$$

→ 오차 계수 K 적용

Method - GS

- Depth mean/covariance/opacity

$$G(\mathbf{x}) = \alpha \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1} (\mathbf{x}-\mu)\right)$$

- 알파 블

$$\mathbf{C} = \sum_{i \in \mathcal{N}} c_i \alpha_i \prod_{j=1}^{i-1} (1-\alpha_j)$$

→ GS에서의 color blending

ure로

$$\mathbf{F}_{\text{BEV}}(\mathbf{x}) = \sum_{i \in \mathcal{G}_{\text{BEV}}} F_i \alpha_i \exp\left(-\frac{1}{2}(\mathbf{x}-\mu_i)^\top \Sigma_i^{-1} (\mathbf{x}-\mu_i)\right)$$

→ GaussianLSS에서의 feature blending

Toward Real-world BEV Perception: Depth Uncertainty Estimation via Gaussian Splatting

CVPR 2025

- **Method**

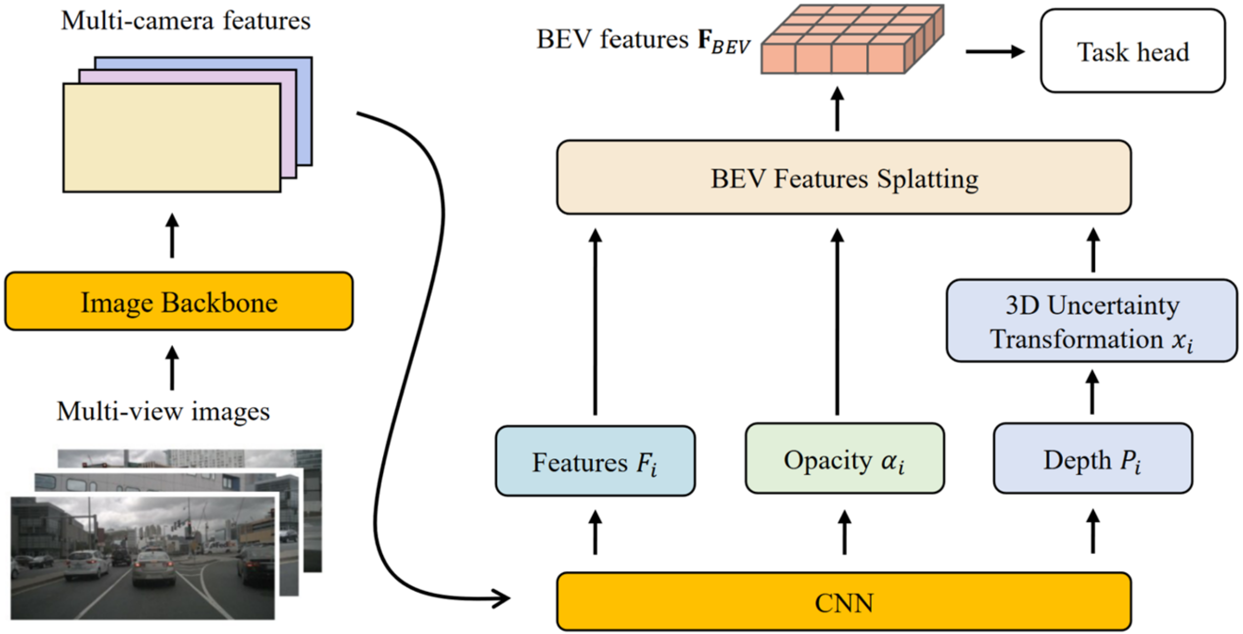


Figure 3. **Overview of GaussianLSS.** Multi-view images are first processed by a backbone network to extract features. They are then input to a simple CNN layer to obtain splat features F_i, opacity α_i, and depth distribution P_i. The predicted depth distribution undergoes an uncertainty transformation to produce a 3D uncertainty x_i. Next, BEV features are obtained through a splatting process, integrating features across views. The resulting BEV features $\mathbf{F}_{\text{BEV}}$, enriched with uncertainty awareness, are used as input to the task-specific head for prediction.

김범준

Toward Real-world BEV Perception: Depth Uncertainty Estimation via Gaussian Splatting

CVPR 2025

- **Experiment**

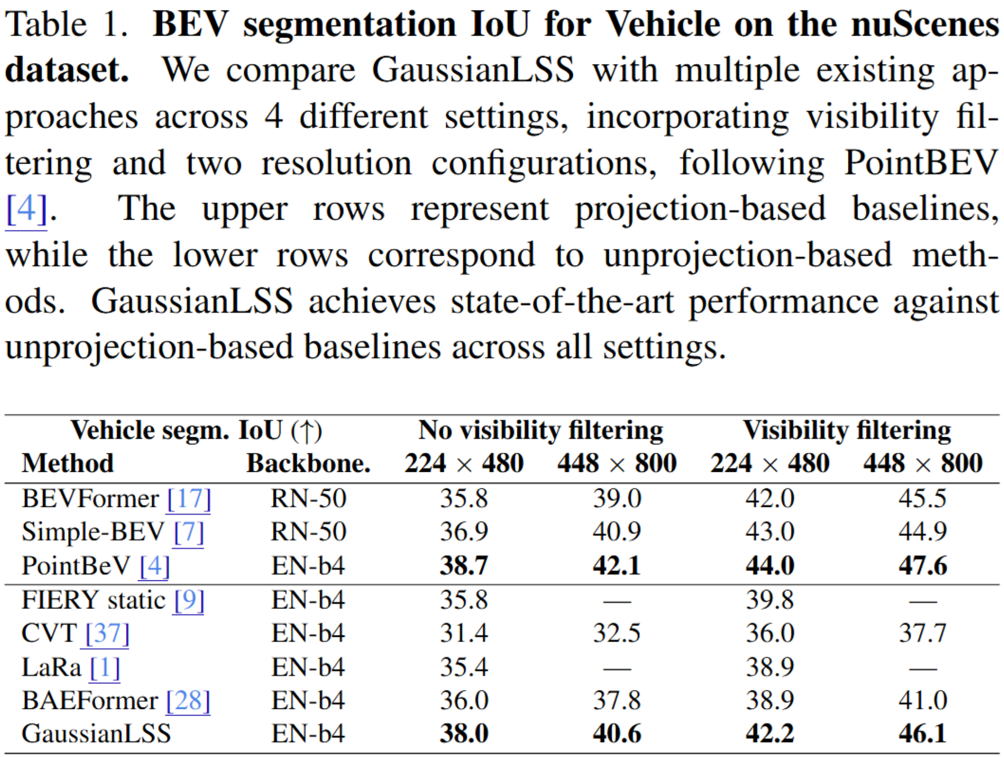


Table 1. **BEV segmentation IoU for Vehicle on the nuScenes dataset.** We compare GaussianLSS with multiple existing approaches across 4 different settings, incorporating visibility filtering and two resolution configurations, following PointBEV [4]. The upper rows represent projection-based baselines, while the lower rows correspond to unprojection-based methods. GaussianLSS achieves state-of-the-art performance against unprojection-based baselines across all settings.

Vehicle segm. IoU (↑)		**No visibility filtering**		**Visibility filtering**	
Method	**Backbone.**	**224 × 480**	**448 × 800**	**224 × 480**	**448 × 800**
BEVFormer [17]	RN-50	35.8	39.0	42.0	45.5
Simple-BEV [7]	RN-50	36.9	40.9	43.0	44.9
PointBeV [4]	EN-b4	**38.7**	**42.1**	**44.0**	**47.6**
FIERY static [9]	EN-b4	35.8	—	39.8	—
CVT [37]	EN-b4	31.4	32.5	36.0	37.7
LaRa [1]	EN-b4	35.4	—	38.9	—
BAEFormer [28]	EN-b4	36.0	37.8	38.9	41.0
GaussianLSS	EN-b4	**38.0**	**40.6**	**42.2**	**46.1**

김범준

Toward Real-world BEV Perception: Depth Uncertainty Estimation via Gaussian Splatting

CVPR 2025

- **Experiment**

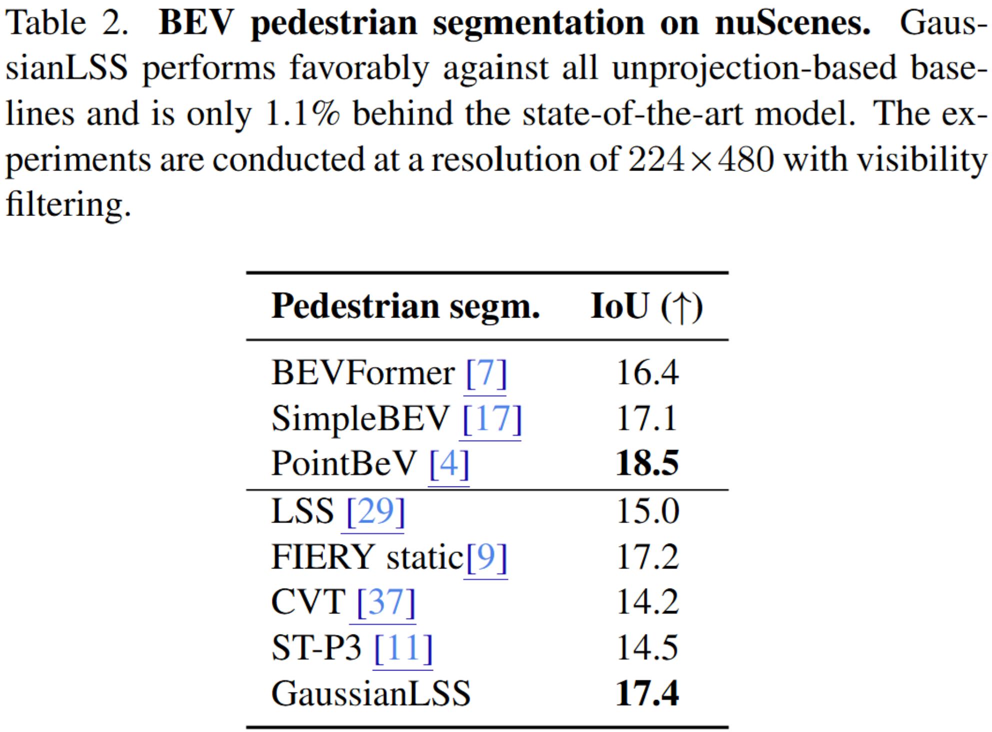


Table 2. **BEV pedestrian segmentation on nuScenes.** GaussianLSS performs favorably against all unprojection-based baselines and is only 1.1% behind the state-of-the-art model. The experiments are conducted at a resolution of 224×480 with visibility filtering.

Pedestrian segm.	**IoU (↑)**
BEVFormer [7]	16.4
SimpleBEV [17]	17.1
PointBeV [4]	**18.5**
LSS [29]	15.0
FIERY static[9]	17.2
CVT [37]	14.2
ST-P3 [11]	14.5
GaussianLSS	**17.4**

Toward Real-world BEV Perception: Depth Uncertainty Estimation via Gaussian Splatting

CVPR 2025

- **Experiment**

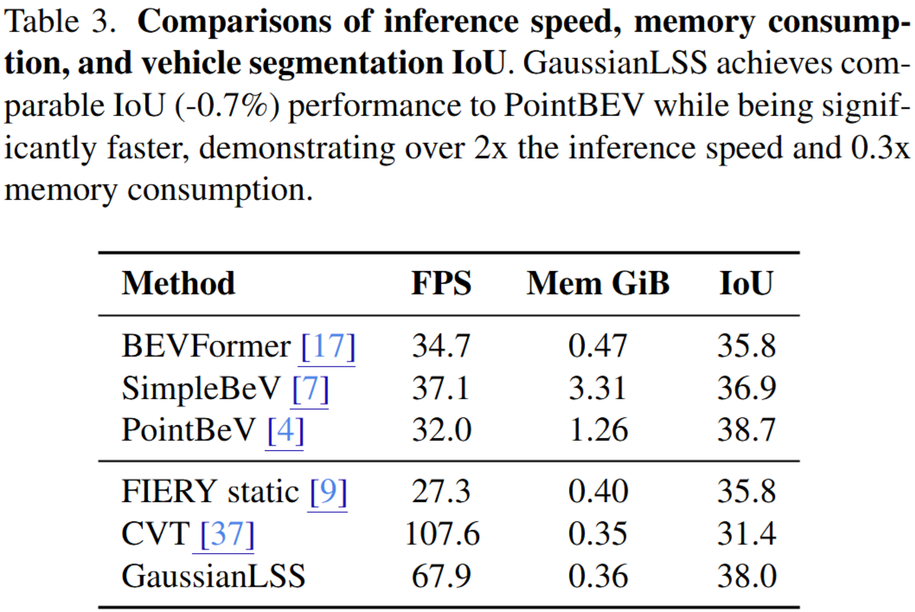


Table 3. **Comparisons of inference speed, memory consumption, and vehicle segmentation IoU**. GaussianLSS achieves comparable IoU (-0.7%) performance to PointBEV while being significantly faster, demonstrating over 2x the inference speed and 0.3x memory consumption.

Method	FPS	Mem GiB	IoU
BEVFormer [17]	34.7	0.47	35.8
SimpleBeV [7]	37.1	3.31	36.9
PointBeV [4]	32.0	1.26	38.7
FIERY static [9]	27.3	0.40	35.8
CVT [37]	107.6	0.35	31.4
GaussianLSS	67.9	0.36	38.0

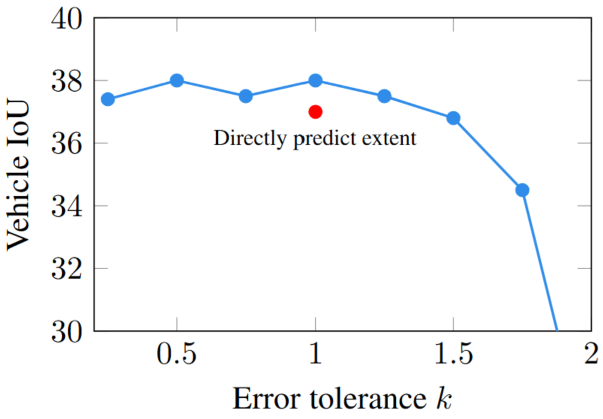


Figure 4. **Sweeping analysis on error tolerance k.** We vary the error tolerance coefficient k across a range of values ($k = [0.25, 2.0]$). The results indicate that performance remains consistent for k values between 0.5 and 1.25. However, when k becomes too large, the IoU drops significantly as the model tolerates excessive ambiguity, causing the features to spread out too much and lose precision. The red dot represents the baseline approach of directly predicting the extent of the 3D mean.

Toward Real-world BEV Perception: Depth Uncertainty Estimation via Gaussian Splatting

CVPR 2025

- ## Experiment

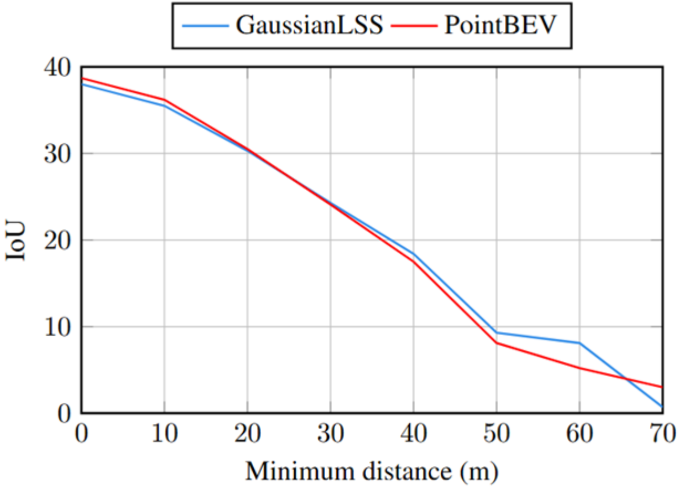


Figure 6. **Impact of distance on vehicle segmentation IoU.** We compare between IoU and distance to the ego-vehicle. Each marker represents the average IoU for vehicles at least d meters away.

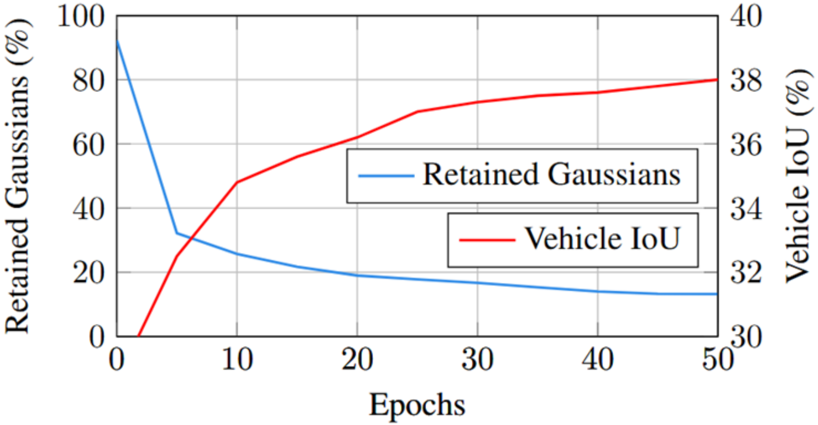


Figure 7. **Proportion of Gaussians retained and Vehicle IoU over training epochs.** The proportion of retained Gaussians($\alpha < 0.01$) reduces to around 20% as the model converges, improving efficiency. Meanwhile, Vehicle IoU steadily increases, showcasing the improved semantic accuracy.

Toward Real-world BEV Perception: Depth Uncertainty Estimation via Gaussian Splatting

CVPR 2025

- **Experiment**

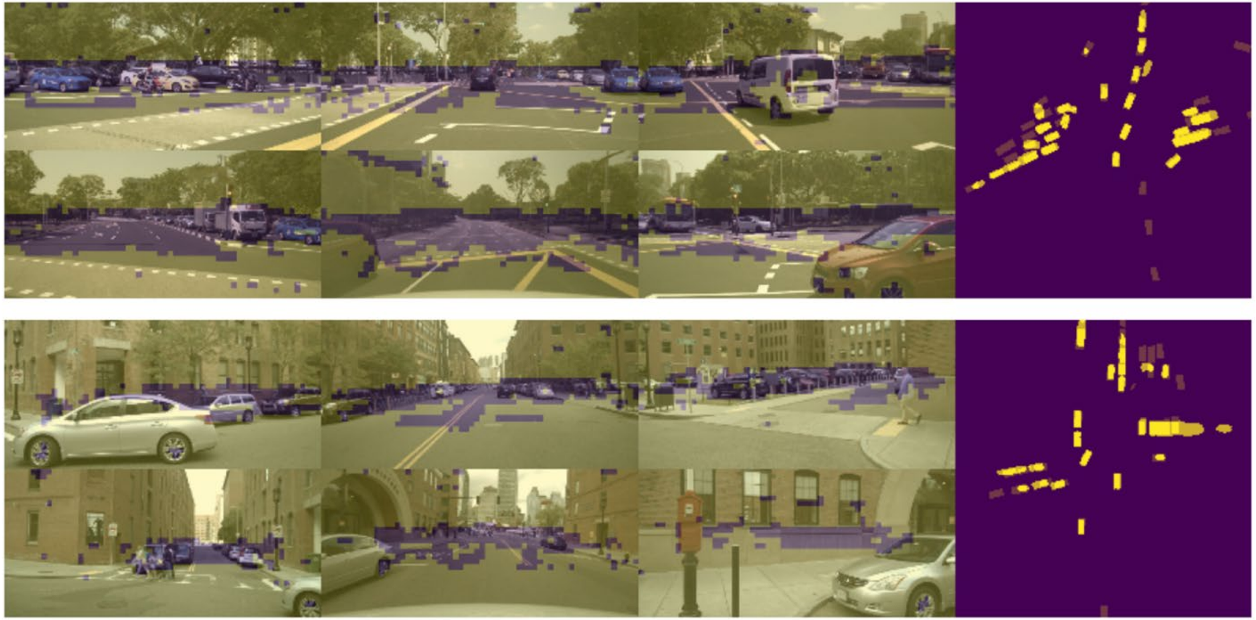


Figure 5. **Qualitative results demonstrating the effectiveness of semantic learning by filtering opacity values below 0.01.** The yellow regions represent masked-out areas during features lifting. The left column shows the six camera views surrounding the ego-vehicle, with the top three views being front-facing and the bottom three being back-facing. The right column depicts BEV predictions overlapped with the ground truth segmentation for reference. The results demonstrate the model's ability to learn meaningful semantic features and accurately project relevant regions to the BEV plane. The ego-vehicle is centered in the map, with visualization highlighting focusing on critical areas.