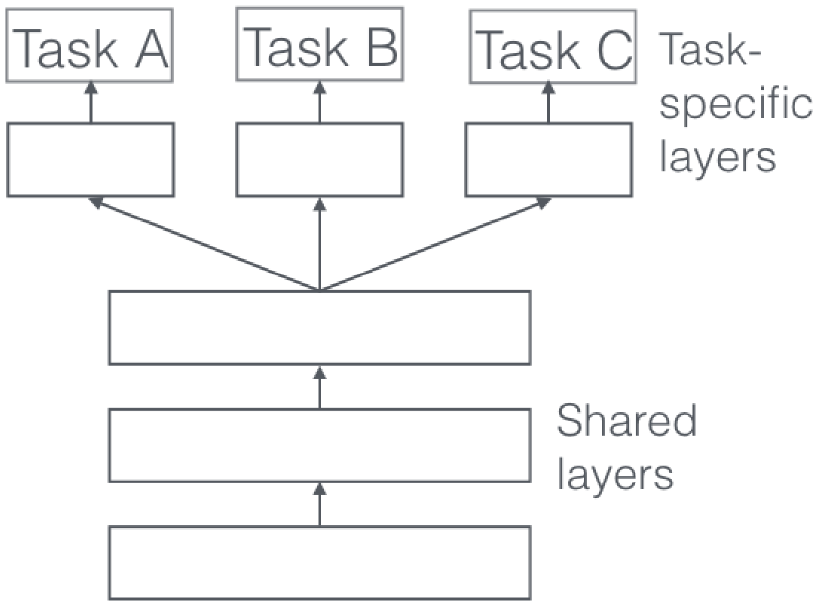
MULTI-TASK LEARNING USING UNCERTAINTY TO WEIGH LOSSES FOR SCENE GEOMETRY AND SEMANTICS

Alex Kendall, Yarin Gal, Roberto Cipolla

CVPR(Spotlight), 2018

Background: Multi-task Learning

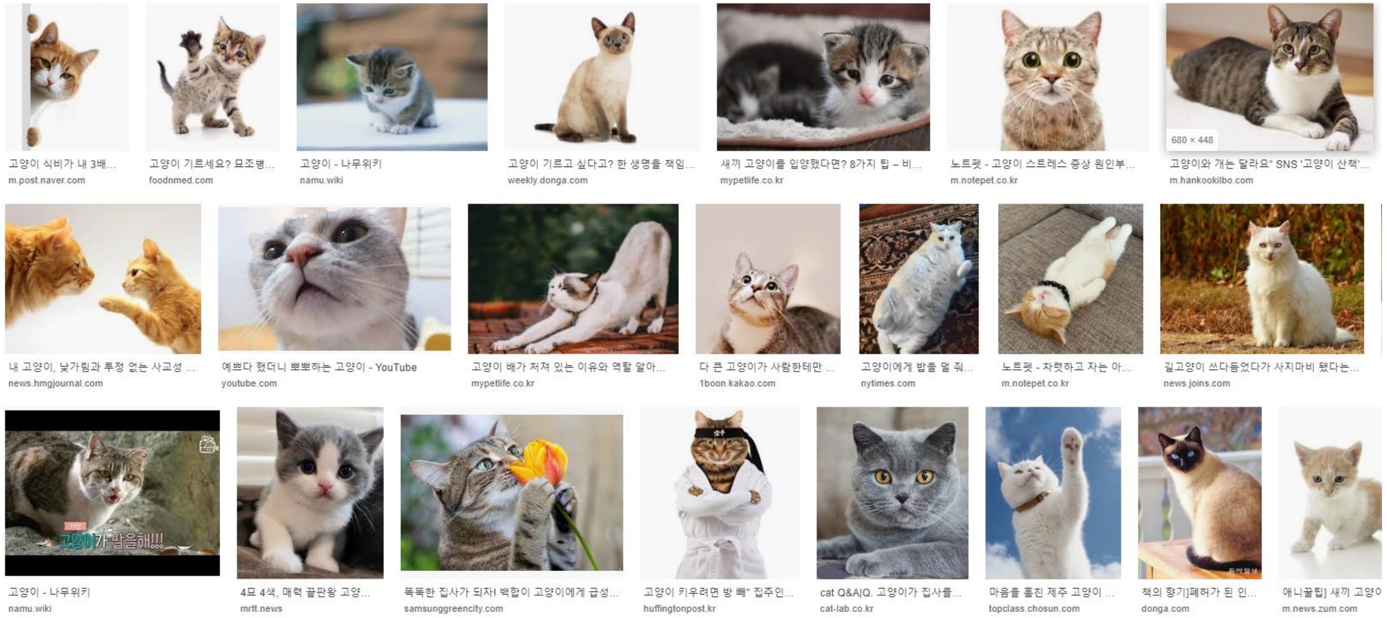
- 하나의 신경망으로 여러가지 작업 동시 진행
  - Ex) Object Detection + Classification
- 개별 학습에 비해 효율적임
- 개별 학습에 비해 더 좋은 성능
  - 작업간의 collaboration?



Background: Uncertainty

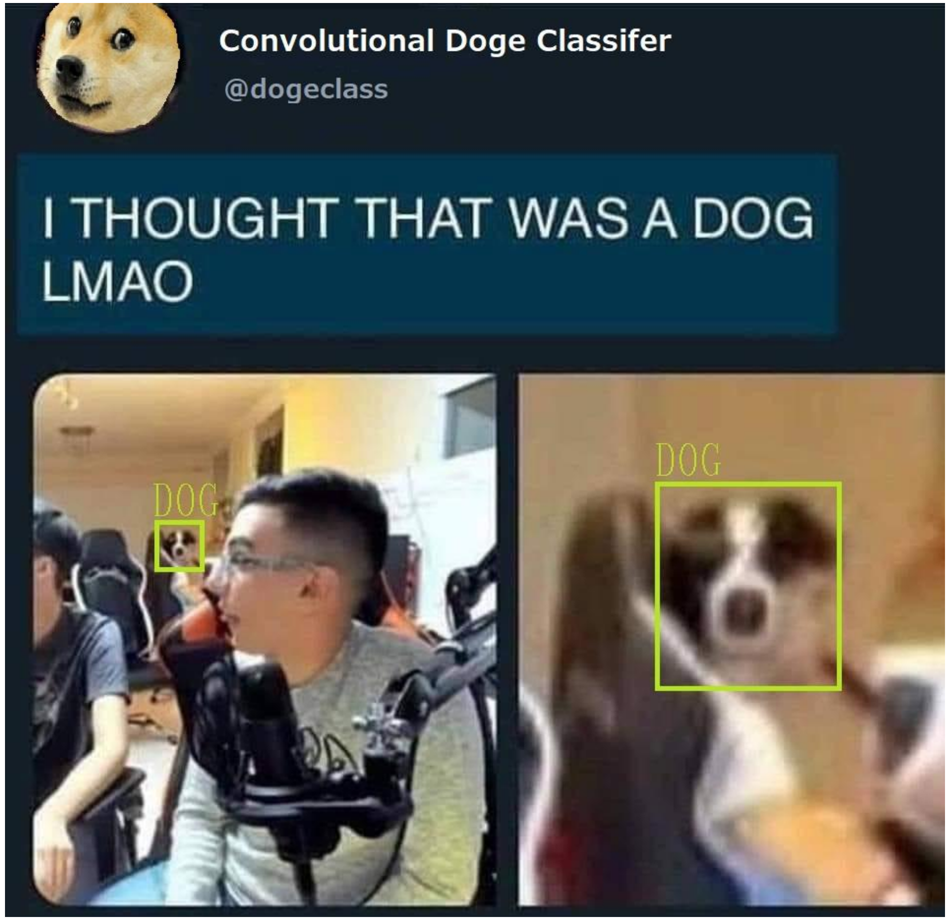
- 고양이와 개 Classifier

Training Data



Input

Background: Uncertainty



Background: Uncertainty

- Epistemic Uncertainty

  (Epistemic : 지식의)

  - 모델 자체의 불확실성
  - 학습 데이터 부족으로 발생
  - 학습 데이터의 증가로 해결
  - 고양이와 개 Classifier

- Aleatoric Uncertainty

  (Aleatoric: 우연의, 도박의)

  - 데이터가 설명 불가능한 정보로 발생
  - 이러한 정보를 설명할 수 있는 변수를 관측하여 해결 가능
  - Data-dependent와 Task-dependent로 나뉨

Alex Kendall and Yarin Gal. 2017. What uncertainties do we need in Bayesian deep learning for computer vision? In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 5580–5590.

Objective

- 다양한 Classification과 Regression을 동시에 학습하는 새로운 Multi-task loss 제안
- Semantic/instance segmentation, depth regression
- 장면(Scene)을 이해하기 위해서는 의미(semantics)와 공간(geometry)의 이해가 필수
- 세 작업을 동시에 진행하여 장면의 의미와 공간 이해 동반

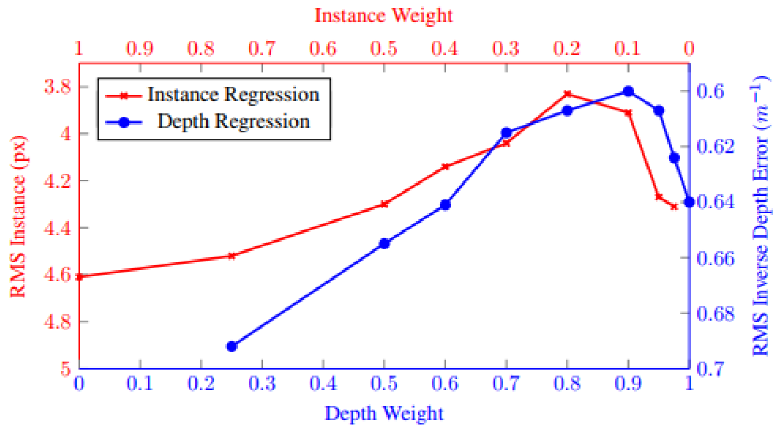
(a) Input image (b) Segmentation output (c) Instance output (d) Depth output

Loss with Uncertainty: Naïve Approach

- Multi-task Learning에서의 기존 방법

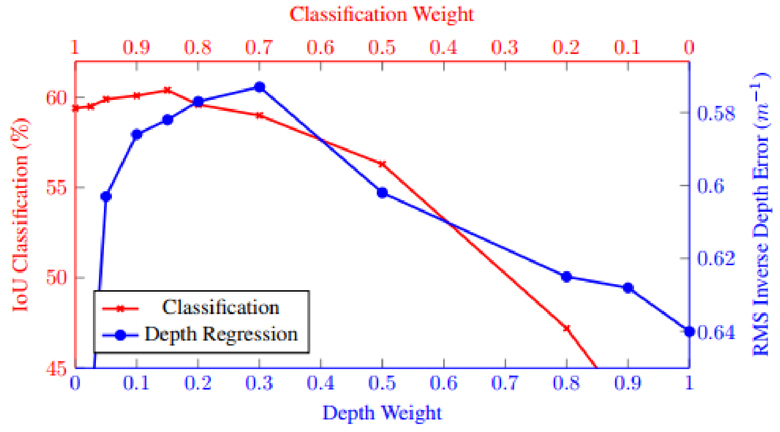
$$L_{total} = \sum_i w_i L_i$$

  - 가중치 w_i에 극도로 예민
  - 탐색 비효율적
  - 좀 더 편한 방법이 필요



Task Weights Instance	Task Weights Depth	Instance Err. [px]	Depth Err. [px]
1.0	0.0	4.61	
0.75	0.25	4.52	0.692
0.5	0.5	4.30	0.655
0.4	0.6	4.14	0.641
0.3	0.7	4.04	0.615
0.2	0.8	3.83	0.607
0.1	0.9	3.91	0.600
0.05	0.95	4.27	0.607
0.025	0.975	4.31	0.624
0.0	1.0		0.640
Learned weights with task uncertainty (this work, Section 3.2)		**3.54**	**0.539**

(b) Comparing loss weightings when learning **instance regression and depth regression**

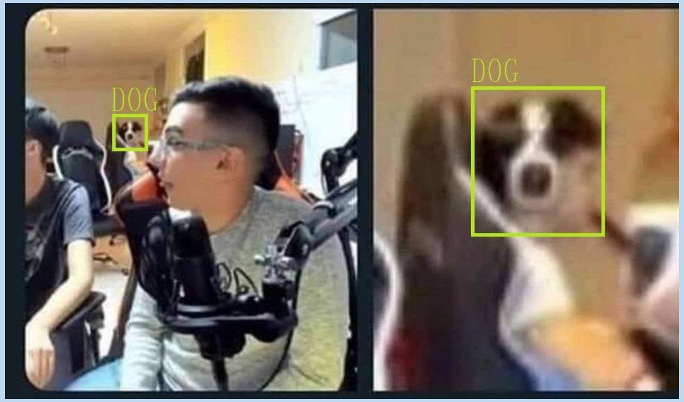


Task Weights Class	Task Weights Depth	Class IoU [%]	Depth Err. [px]
1.0	0.0	59.4	-
0.975	0.025	59.5	0.664
0.95	0.05	59.9	0.603
0.9	0.1	60.1	0.586
0.85	0.15	60.4	0.582
0.8	0.2	59.6	0.577
0.7	0.3	59.0	0.573
0.5	0.5	56.3	0.602
0.2	0.8	47.2	0.625
0.1	0.9	42.7	0.628
0.0	1.0	-	0.640
Learned weights with task uncertainty (this work, Section 3.2)		**62.7**	**0.533**

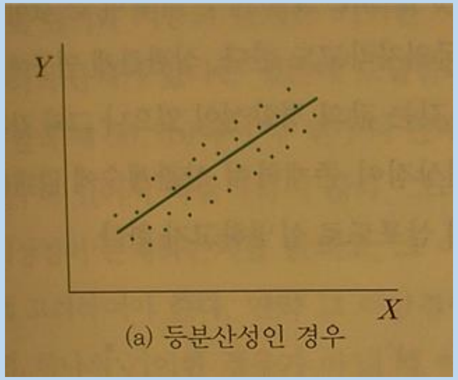
(a) Comparing loss weightings when learning **semantic classification and depth regression**

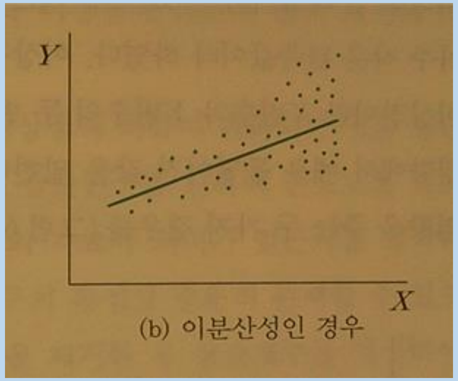
Loss with Uncertainty: Aleatoric

- Data-dependent
  - 입력 데이터에 의존적



- Task-dependent
  - 입력 데이터에 비의존적
  - 모든 입력 데이터에 공통적으로 존재, 비슷한 값을 가짐
    - Homoscedastic (등분산성)
  - Task 별로 다름





Multi-task Likelihood

- x: 입력 데이터, f^W(x) : NN의 출력, y: 타겟
- Regression tasks Likelihood :

$$p(\mathbf{y}|\mathbf{f}^{\mathbf{W}}(\mathbf{x})) = \mathcal{N}(\mathbf{f}^{\mathbf{W}}(\mathbf{x}), \sigma^2) \tag{2}$$

- Classification tasks Likelihood :

$$p(\mathbf{y}|\mathbf{f}^{\mathbf{W}}(\mathbf{x})) = \mathrm{Softmax}(\mathbf{f}^{\mathbf{W}}(\mathbf{x})). \tag{3}$$

- In sufficient statistics, Multi-task Likelihood :

$$p(\mathbf{y}_1, ..., \mathbf{y}_K|\mathbf{f}^{\mathbf{W}}(\mathbf{x})) = p(\mathbf{y}_1|\mathbf{f}^{\mathbf{W}}(\mathbf{x}))...p(\mathbf{y}_K|\mathbf{f}^{\mathbf{W}}(\mathbf{x})) \tag{4}$$

- Sufficient Statistics (충분통계량)

모수 자체는 아니지만 모수를 추정할 수 있는 충분한 정보를 가진 값

Multi-task Likelihood

1. Regression + Regression

$$\log p(\mathbf{y}|\mathbf{f}^{\mathbf{W}}(\mathbf{x})) \propto -\frac{1}{2\sigma^2}||\mathbf{y} - \mathbf{f}^{\mathbf{W}}(\mathbf{x})||^2 - \log\sigma \quad (5)$$

$$\begin{aligned} p(\mathbf{y}_1, \mathbf{y}_2|\mathbf{f}^{\mathbf{W}}(\mathbf{x})) &= p(\mathbf{y}_1|\mathbf{f}^{\mathbf{W}}(\mathbf{x})) \cdot p(\mathbf{y}_2|\mathbf{f}^{\mathbf{W}}(\mathbf{x})) \\ &= \mathcal{N}(\mathbf{y}_1; \mathbf{f}^{\mathbf{W}}(\mathbf{x}), \sigma_1^2) \cdot \mathcal{N}(\mathbf{y}_2; \mathbf{f}^{\mathbf{W}}(\mathbf{x}), \sigma_2^2). \end{aligned} \quad (6)$$

$$\begin{aligned} \mathcal{L}(\mathbf{W}, \sigma_1, \sigma_2) &= -\log p(\mathbf{y}_1, \mathbf{y}_2|\mathbf{f}^{\mathbf{W}}(\mathbf{x})) \\ &\propto \frac{1}{2\sigma_1^2}||\mathbf{y}_1 - \mathbf{f}^{\mathbf{W}}(\mathbf{x})||^2 + \frac{1}{2\sigma_2^2}||\mathbf{y}_2 - \mathbf{f}^{\mathbf{W}}(\mathbf{x})||^2 + \log\sigma_1\sigma_2 \\ &= \underbrace{\frac{1}{2\sigma_1^2}\mathcal{L}_1(\mathbf{W}) + \frac{1}{2\sigma_2^2}\mathcal{L}_2(\mathbf{W}) + \log\sigma_1\sigma_2}_{\text{미분가능}} \end{aligned} \quad (7)$$

Multi-task Likelihood

2. Regression + Classification

$$p(\mathbf{y}|\mathbf{f}^{\mathbf{W}}(\mathbf{x}),\sigma) = \text{Softmax}(\frac{1}{\sigma^2}\mathbf{f}^{\mathbf{W}}(\mathbf{x})) \qquad (8)$$

$$\log p(\mathbf{y}=c|\mathbf{f}^{\mathbf{W}}(\mathbf{x}),\sigma) = \frac{1}{\sigma^2}f_c^{\mathbf{W}}(\mathbf{x}) - \log\sum_{c'}\exp\left(\frac{1}{\sigma^2}f_{c'}^{\mathbf{W}}(\mathbf{x})\right) \qquad (9)$$

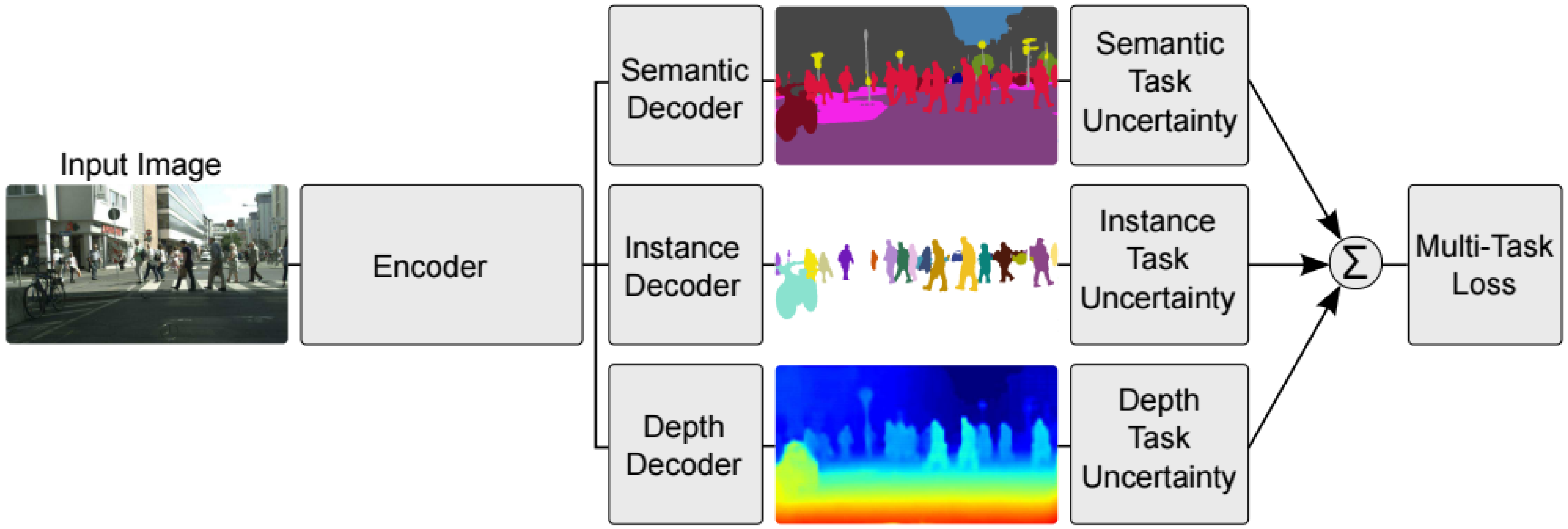
$$\begin{aligned}
\mathcal{L}(\mathbf{W},\sigma_1,\sigma_2) &= -\log p(\mathbf{y}_1,\mathbf{y}_2=c|\mathbf{f}^{\mathbf{W}}(\mathbf{x})) \\
&= -\log \mathcal{N}(\mathbf{y}_1;\mathbf{f}^{\mathbf{W}}(\mathbf{x}),\sigma_1^2)\cdot \text{Softmax}(\mathbf{y}_2=c;\mathbf{f}^{\mathbf{W}}(\mathbf{x}),\sigma_2) \\
&= \frac{1}{2\sigma_1^2}||\mathbf{y}_1-\mathbf{f}^{\mathbf{W}}(\mathbf{x})||^2 + \log\sigma_1 - \log p(\mathbf{y}_2=c|\mathbf{f}^{\mathbf{W}}(\mathbf{x}),\sigma_2) \\
&= \frac{1}{2\sigma_1^2}\mathcal{L}_1(\mathbf{W}) + \frac{1}{\sigma_2^2}\mathcal{L}_2(\mathbf{W}) + \log\sigma_1 \\
&\quad + \log\frac{\sum_{c'}\exp\left(\frac{1}{\sigma_2^2}f_{c'}^{\mathbf{W}}(\mathbf{x})\right)}{\left(\sum_{c'}\exp\left(f_{c'}^{\mathbf{W}}(\mathbf{x})\right)\right)^{\frac{1}{\sigma_2^2}}} \\
&\approx \underline{\frac{1}{2\sigma_1^2}\mathcal{L}_1(\mathbf{W}) + \frac{1}{\sigma_2^2}\mathcal{L}_2(\mathbf{W}) + \log\sigma_1 + \log\sigma_2,}
\end{aligned} \qquad (10)$$

when σ -> 1

미분가능

Scene Understanding Model

- Deep Convolutional Encoder Decoder 사용
- Task 별 Decoder 분리
- 네트워크의 마지막 층에서 Task 별 Task Uncertainty 분산 예측
- Multi-task loss로부터 학습



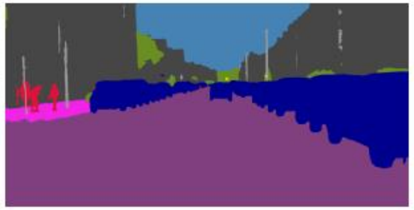
Scene Understanding Model

- Semantic Segmentation
  - cross-entropy loss 사용하여 pixel단위 class 확률을 학습
- Depth Regression
  - L1 loss 사용하여 pixel단위 inverse depth를 학습
  - 하늘과 같이 깊이가 무한대인 점을 표현하기 위해 inverse depth 사용

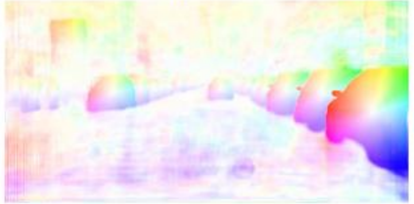
Scene Understanding Model

- Instance Segmentation
  - L1 loss 사용하여 각 pixel로부터 Instance의 중심을 가리키는 벡터 학습
  - 이를 OPTICS 알고리즘으로 clustering
  - Instance가 가려져 있어도 구분 가능(우측 하단)
    - semantics와 geometry 정보의 영향

(a) Input Image



(b) Semantic Segmentation



(c) Instance vector regression

(d) Instance Segmentation

(a) Input Image

(b) Instance Segmentation

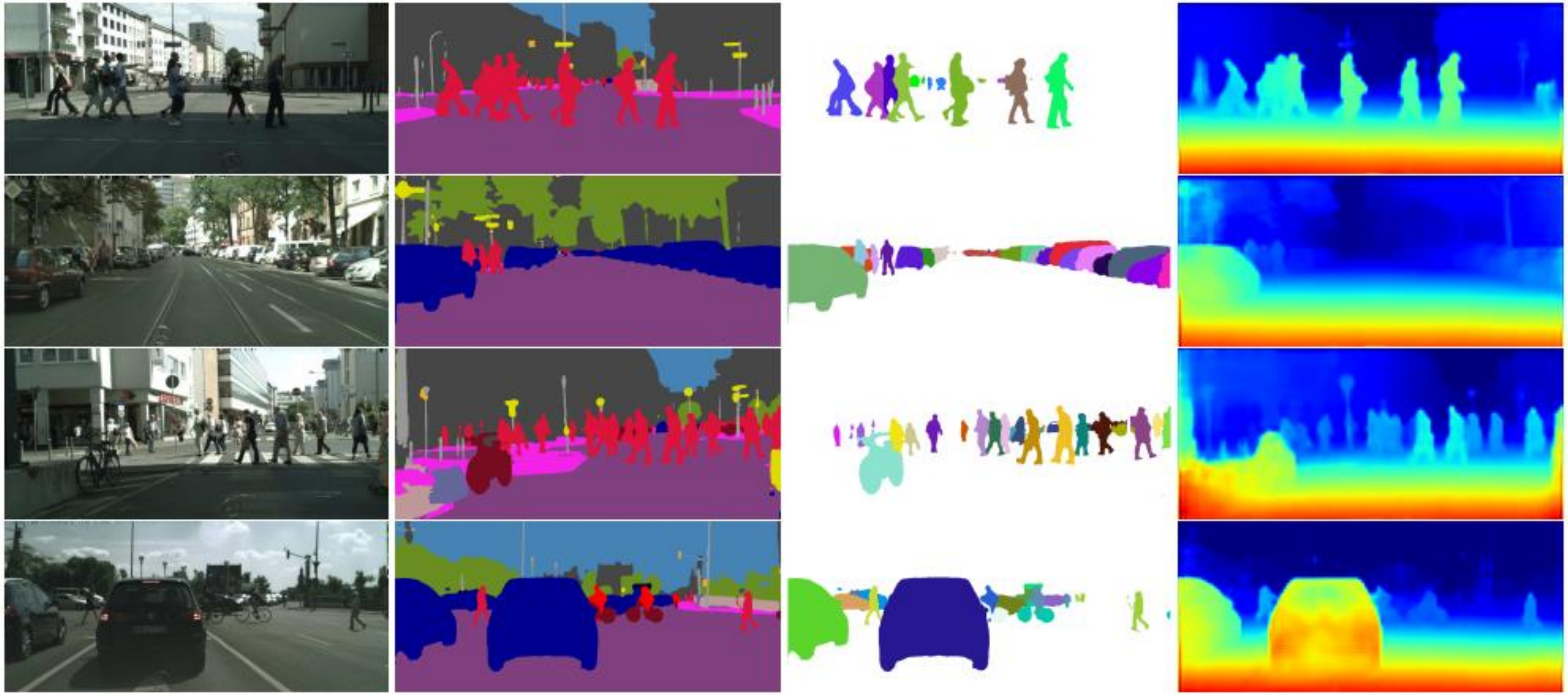
Experiments

Loss	Task Weights Seg.	Inst.	Depth	Segmentation IoU [%]	Instance Mean Error [px]	Inverse Depth Mean Error [px]
Segmentation only	1	0	0	59.4%	-	-
Instance only	0	1	0	-	4.61	-
Depth only	0	0	1	-	-	0.640
Unweighted sum of losses	0.333	0.333	0.333	50.1%	3.79	0.592
Approx. optimal weights	0.89	0.01	0.1	62.8%	3.61	0.549
2 task uncertainty weighting	✓	✓		61.0%	**3.42**	-
2 task uncertainty weighting	✓		✓	62.7%	-	0.533
2 task uncertainty weighting		✓	✓	-	3.54	0.539
3 task uncertainty weighting	✓	✓	✓	**63.4%**	3.50	**0.522**

Experiments

Method	Semantic Segmentation				Instance Segmentation				Monocular Disparity Estimation	
	IoU class	iIoU class	IoU cat	iIoU cat	AP	AP 50%	AP 100m	AP 50m	Mean Error [px]	RMS Error [px]
Semantic segmentation, instance segmentation and depth regression methods (this work)										
Multi-Task Learning	78.5	57.4	89.9	77.7	19.0	35.9	30.9	33.0	2.92	5.88
Semantic segmentation and instance segmentation methods										
Uhrig et al. [41]	64.3	41.6	85.9	73.9	8.9	21.1	15.3	16.7	-	-
Instance segmentation only methods										
Mask R-CNN [19]	-	-	-	-	26.2	49.9	37.6	40.1	-	-
Deep Watershed [4]	-	-	-	-	19.4	35.3	31.4	36.8	-	-
R-CNN + MCG [13]	-	-	-	-	4.6	12.9	7.7	10.3	-	-
Semantic segmentation only methods										
DeepLab V3 [10]	81.3	60.9	91.6	81.7	-	-	-	-	-	-
PSPNet [44]	81.2	59.6	91.2	79.2	-	-	-	-	-	-
Adelaide [31]	71.6	51.7	87.3	74.1	-	-	-	-	-	-

Experiments



(a) Input image (b) Segmentation output (c) Instance output (d) Depth output