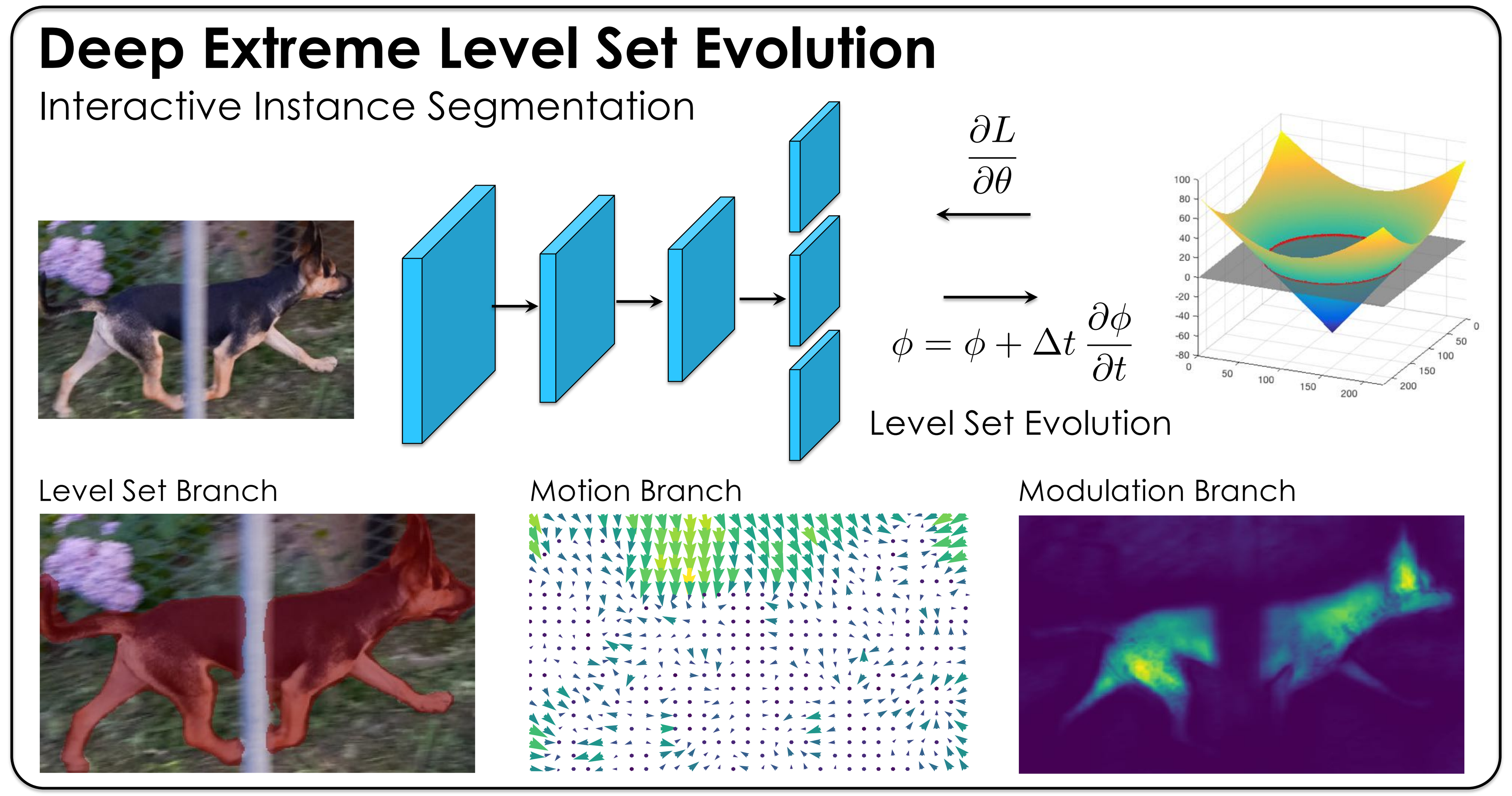
Object Instance Annotation with Deep Extreme Level Set Evolution

Zian Wang[1], David Acuna*[2,3,4], Huan Ling*[2,3], Amlan Kar[2,3], Sanja Fidler[2,3,4]
Tsinghua University[1], University of Toronto[2], Vector Institute[3], NVIDIA[4]

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

Deep Extreme Level Set Evolution



DELSE combines powerful CNN image feature extraction with Level Set Evolution. It is end-to-end differentiable, and produces "well behaved" object contours.

Level Set Formulation

Level Set Representation

- Implicit curve with level set function ϕ $\quad \mathcal{C} = \{(x,y) | \phi(x,y) = 0\}$

  Foreground: $\{(x,y) \in \Omega_I | \phi(x,y) > 0\}$ Background: $\{(x,y) \in \Omega_I | \phi(x,y) < 0\}$

- Curve evolution with level sets

$$\frac{\partial C(s,t)}{\partial t} = V\vec{N} \Leftrightarrow \frac{\partial \phi}{\partial t} = -V|\nabla \phi|$$

$$\phi_{i+1}(x,y) = \phi_i(x,y) + \Delta t \frac{\partial \phi_i}{\partial t}$$

Level Set Energy Design

- ***Motion Term***: determines the motion of level set evolution. DELSE predicts a vector field $\vec{V}_\theta$ with *motion branch* and evolve with

$$\left[\frac{\partial \phi_i}{\partial t}\right]_{\text{motion}} = -\langle \vec{V}_\theta, \nabla \phi_i \rangle$$

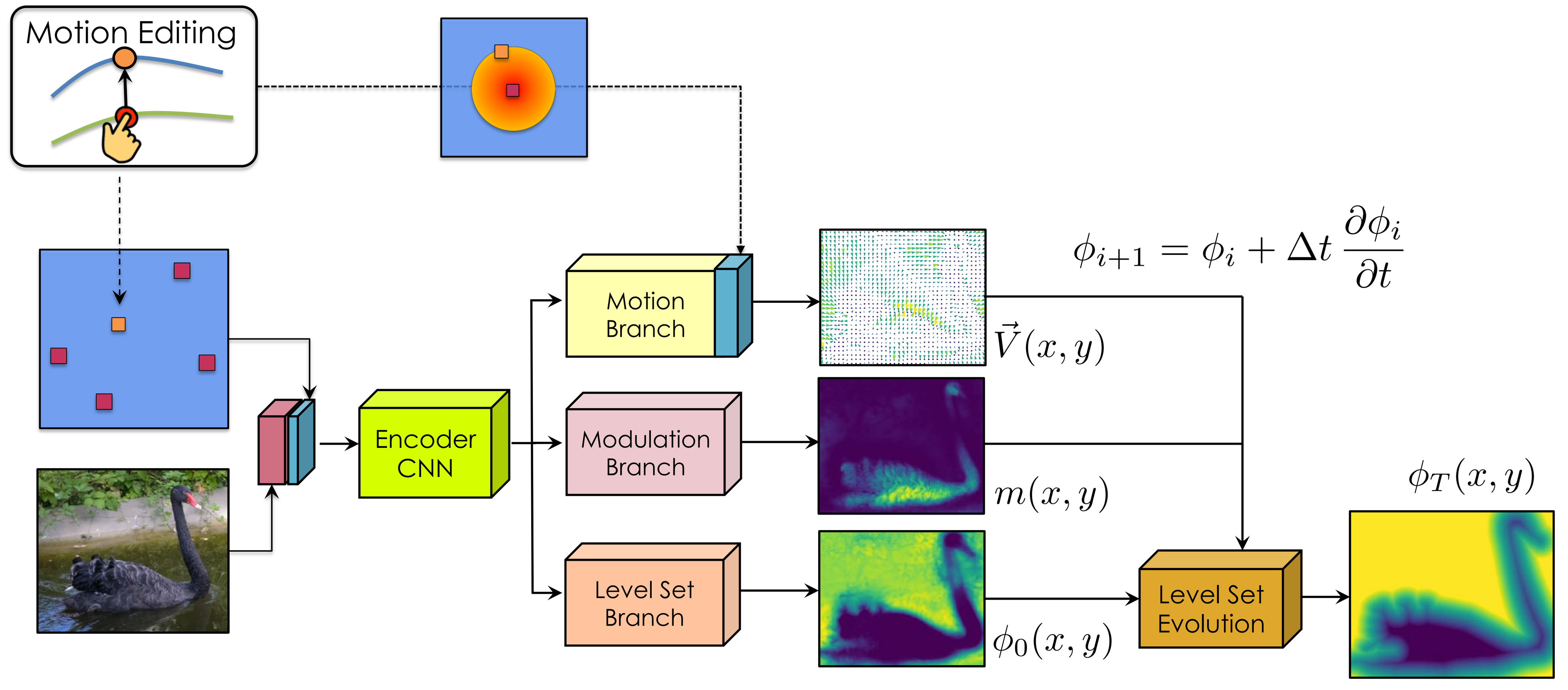
- ***Curvature Term***: To make the curve's shape generally well behaved, DELSE regularize the predicted curve by moving it in the direction of its curvature. This term is selective with a learned modulation function m_θ.

$$\left[\frac{\partial \phi_i}{\partial t}\right]_{\text{curvature}} = m_\theta \kappa |\nabla \phi_i| = m_\theta |\nabla \phi_i| \operatorname{div}\left(\frac{\nabla \phi_i}{|\nabla \phi_i|}\right)$$

- ***Regularization Term***: To maintain a desirable shape of LSF, DELSE regularize $|\nabla \phi|$ to be either close to 0 or 1 with

$$\left[\frac{\partial \phi_i}{\partial t}\right]_{\text{reg}} = \operatorname{div}\left(p'(|\nabla \phi_i|)\frac{\nabla \phi_i}{|\nabla \phi_i|}\right)$$

Model Architecture



Architecture of DELSE: Extreme points are encoded as a heat map and concatenated with the image, which are then passed to the encoder CNN. A multi-branch architecture is used to predict the initial curve and parameters used in level set evolution.

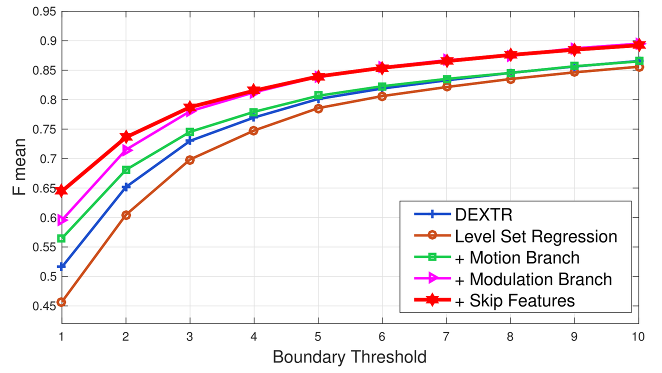
Results

Model	Bicycle	Bus	Person	Train	Truck	Motorcycle	Car	Rider	**mIoU**	F mean(1 pix)	F mean(2 pix)
DEXTR*	71.92	87.42	78.36	78.11	84.88	72.41	84.62	75.18	79.11	54.00	68.60
DELSE*	**74.32**	**88.85**	**80.14**	**80.35**	**86.05**	**74.10**	**86.35**	**76.74**	**80.86**	**60.29**	**74.40**
DEXTR [29]	76.36	88.58	82.44	76.40	87.53	75.20	87.17	79.06	81.59	60.65	73.85
Level Set Regression	76.05	88.21	82.40	78.69	86.50	74.31	87.17	78.99	81.54	58.87	72.08
DELSE	**77.83**	**89.56**	**83.42**	**82.45**	**88.11**	**77.16**	**88.29**	**79.98**	**83.35**	**64.35**	**77.62**

Quantitative Evaluation on Cityscapes. mIoU for region similarity and F metric for boundary.

Model	J mean	J recall	F mean	F recall
DEXTR	82.4	94.2	84.5	93.5
Level Set Regression	81.7	90.9	83.4	91.4
+ Motion Term	84.0	94.9	84.7	94.0
+ Modulation Term	84.8	95.0	87.5	**95.1**
+ Skip Features	**85.6**	**95.1**	**87.8**	94.8

Ablation study on DAVIS.



Multi-scale boundary evaluation on DAVIS.

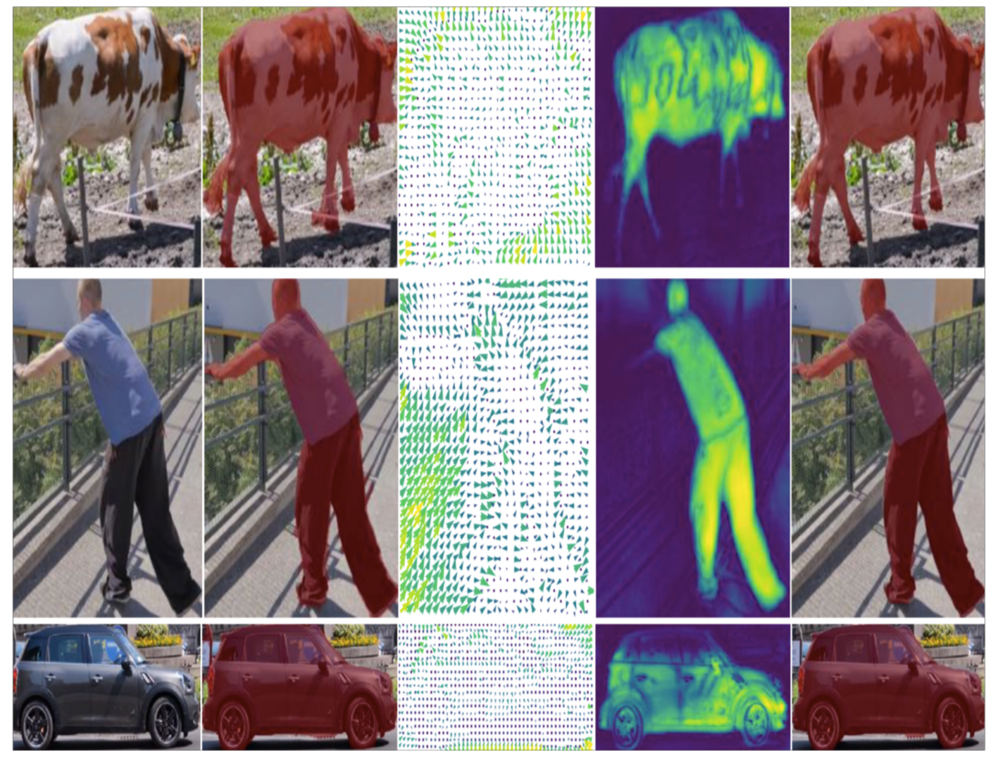
Model	mIOU	F mean
DELSE (Full data)	83.35	77.62
DELSE* (10 of 16 cities)	82.45	75.85
Motion Editing Clicks	**mIOU**	**F mean**
1	**84.73**	**79.64**
2	**85.97**	**81.34**
3	**86.83**	**82.52**
Extreme Points Clicks	**mIOU**	**F mean**
1	83.60	78.27
2	84.49	79.67
3	84.94	80.53

Interactive correction on Cityscapes. Corrections are used with DELSE*, which is trained on 10 out of 16 cities.

Qualitative Results

Qualitative results on Cityscapes. Note that our model takes ground-truth boxes as input, following the setting of Polygon-RNN.

Qualitative results for occluded objects on Cityscapes. **Top row**: ground-truth, **Bottom row**: DELSE



Visualization of CNN branches outputs.

Visualization of Level Set Evolution through time.

We gratefully acknowledge support from Vector Institute and NVIDIA for computational resources.