Joint Coarse-and-Fine Reasoning For Deep Optical Flow

Victor Vaquero, German Ros
Francesc Moreno-Noguer, Antonio M. Lopez, Alberto Sanfeliu
Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Barcelona, Spain
Universitat Autonoma de Barcelona, Campus UAB, Barcelona, Spain
Computer Vision Center, Campus UAB, Barcelona, Spain
Toyota Research Institute, Palo Alto, USA

Institut de Robòtica i Informàtica Industrial

CSIC Spanish Council of Research

UPC UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

TOYOTA RESEARCH INSTITUTE

UAB Universitat Autònoma de Barcelona

CVC Centre de Visió per Computador

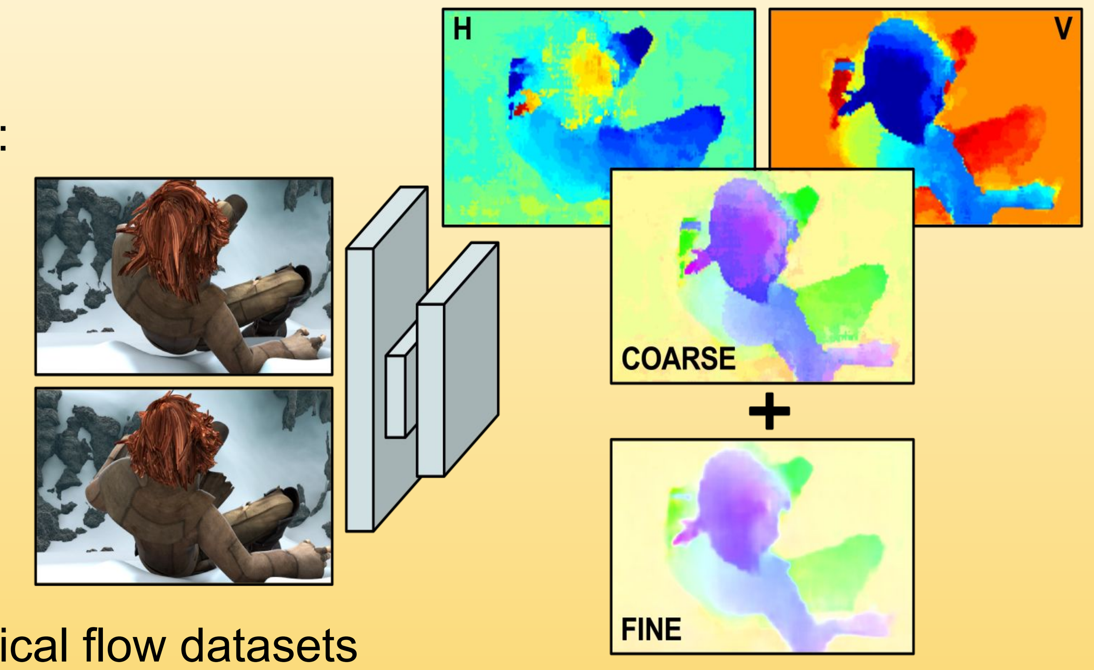
General Idea & Fundamentals

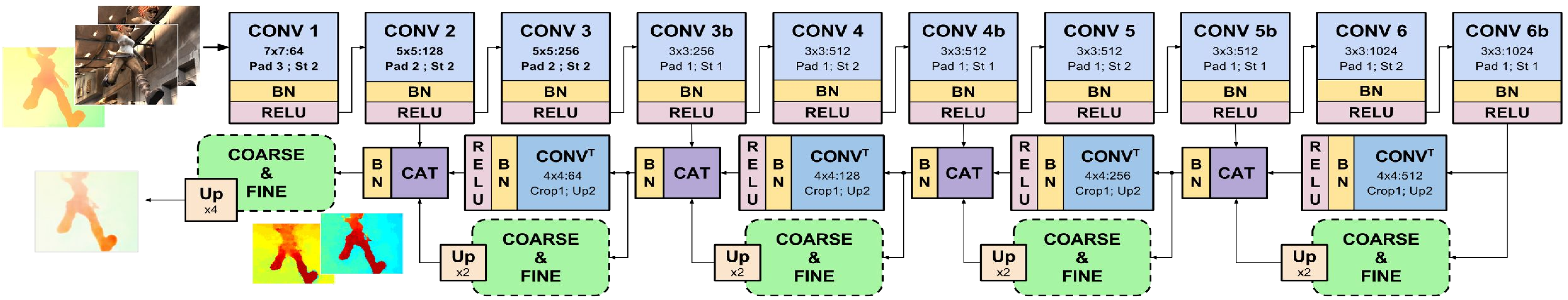
- Typically CNN-Based learning solutions: Classification [1] or Regression tasks [2]
- We present a novel joint **Coarse-and-Fine** reasoning for dense pixel-wise estimation tasks:

  Coarse general rough solution over a discrete classification space

  Fine details of the solution are obtained over a continuous regression space

- We prove that estimate both components jointly is beneficial for improving accuracy
- Our architecture treats the fine estimation as a refinement built on top of the coarse one
- We apply our approach to the optical flow estimation challenging problem and validate it against state-of-the-art CNN-based solutions trained from scratch and tested on large optical flow datasets
- First totally CNN-based optical flow approach introduced in FlowNet [2]. Updated in FlowNet-2 [3]
- Residual blocks has proven to yield a notorious improvement in speed and accuracy [4]





Approach

- We define a basic Optical Flow architecture as combination of blocks $G_\theta(\cdot)$ and $F_\theta(\cdot)$
  - Initially, for an RGB image $\mathcal{X} \in \mathbb{R}^{HxWx3}$ ➡ $F_\theta(\mathcal{X})$, extracts features from the image, based on [2]
  - Then, $G_\theta(F_\theta(\mathcal{X})) = \hat{y} \in \mathbb{R}^{HxWx2}$ transforms these features into Optical Flow predictions.
- Our $G_\theta(\cdot)$ generates to branched predictions: $\hat{y}^{reg}$ & $\hat{y}^{class}$ solving respectively the regression and classification tasks.
- Final Coarse-and-Fine Loss: $\mathcal{L}_{CaF}(\hat{y}, y) = \mathcal{L}_{coarse}(\hat{y}^{class}, y^{class}) + \lambda \mathcal{L}_{fine}(\hat{y}^{reg}, y^{reg})$

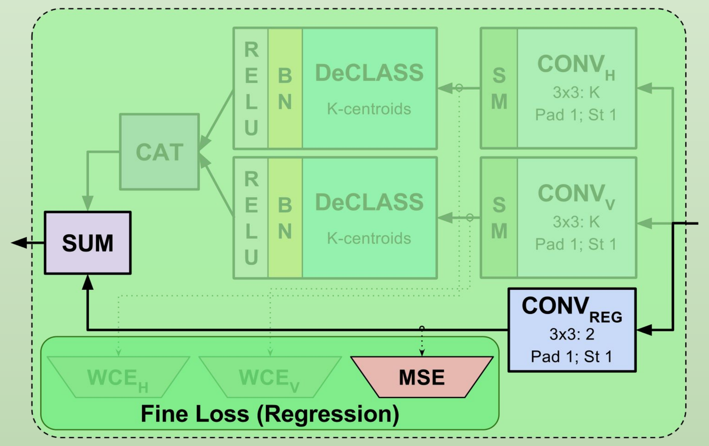
Classification Problem

- K classes: $$I_k = \begin{cases} (-\infty,\ C_1 + \delta/2), & if\ \ k = 1 \\ [C_k - \delta/2,\ C_k + \delta/2), & if\ 1 < k < K \\ [C_K - \delta/2, +\infty), & if\ k = K \end{cases}$$
- With centroids: $C_k = m_r + \delta(k-1),\ \ k \in 1,\dots,K$
- Weighted Cross Entropy Loss:

$$\mathcal{L}_{WCE} = -\sum_{i,j,k}^{H,W,K} \omega(y_{i,j}^{class})\, Id_{[y_{i,j}^{class}]}(log(\hat{y}_{i,j,k}^{class}))$$

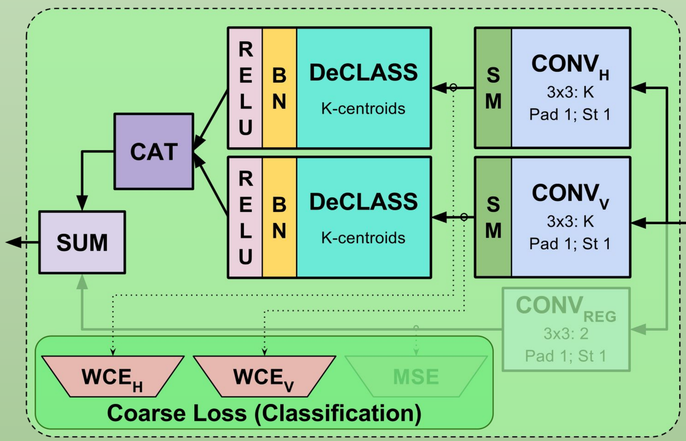
Experiments & Results

Three Different experimental configurations:



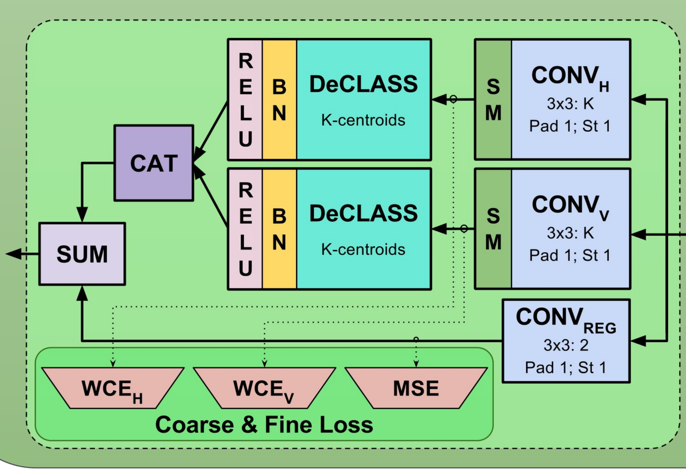
Regression Baseline:

- Regression-Only loss (Mean Square Error)
- Explicit classification part is deactivated
- FlowNet like architecture



Classification Baseline:

- Classification-Only loss (Weighted Cross Entropy)
- Explicit regression part is deactivated
- Horizontal and Vertical classification decomposition



Full Approach:

- Combined Coarse-and-Fine loss activated
- Experimented with {5,21,41} classes
- Up to 16% of improvement wrt baselines

	F.Chairs Validation	Sintel Train Clean	Sintel Train Final	Sintel Test Clean	Sintel Test Final
Regression	3.78 (100)	6.93 (100)	7.66 (100)	9.98 (100)	10.72 (100)
Class-5c	6.99 (184.7)	9.66 (139.4)	10.20 (133.1)	13.11 (131.3)	13.54 (126.3)
Class-21c	4.06 (107.3)	7.91 (114.1)	8.50 (110.9)	10.70 (107.1)	11.34 (105.8)
Class-41c	3.81 (100.7)	7.69 (110.87)	8.38 (109.3)	10.66 (106.7)	11.53 (107.5)
CaF-5c	3.55 (93.8)	6.85 (98.8)	7.54 (98.5)	9.98 (99.9)	10.69 (99.7)
CaF-21c	3.44 (90.9)	6.76 (97.5)	7.43 (96.9)	9.88 (98.9)	10.53 (98.2)
CaF-41c	3.47 (91.7)	6.75 (97.4)	7.39 (96.4)	9.77 (97.8)	10.48 (97.7)
CaF-Full-5c	3.25 (85.8)	6.85 (98.84)	7.72 (100.7)	9.74 (97.5)	10.51 (98.1)
CaF-Full-21c	3.23 (85.3)	6.75 (97.34)	7.59 (99.0)	9.57 (95.8)	10.28 (95.9)
CaF-Full-41c	3.18 (84.0)	6.51 (93.84)	7.28 (95.0)	9.42 (94.3)	10.18 (95.0)

Bibliography:

[1] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in The IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2015.

[2] A. Dosovitskiy, P. Fischer, E. Ilg, P. Caner Hazirbas, V. Golkov, P. van der Smagt, D. Cremers, and T.Brox, "Flownet: Learning optical flow with convolutional networks," in The IEEE International Conference on Computer Vision (ICCV), 2015.

[3] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy and T. Brox, "FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks", in The IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2017.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in The IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2016.

Acknowledgements:

This work was partially supported by European AEROARMS project (H2020-ICT-2014-1-644271) and CICYT projects ColRobTransp (DPI2016-78957-R), ROBINSTRUCT(TIN2014-58178-R). Authors thank Nvidia for GPU hardware donation.