Convolutional STN for Weakly Supervised Object Localization

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

Université du Québec



LABORATOIRE D'IMAGERIE, DE VISION ET D'INTELLIGENCE ARTFICIELLE Akhil Meethal Marco Pedersoli Soufiane Belharbi Eric Granger

Overview

- Weakly supervised Object Localization(WSOL)
- □ Convolutional STN(CSTN)
- Empirical Study
- **Conclusion**

Weakly supervised Object Localization(WSOL)



Why WSOL?



Annotation time for different types of supervision[Bilen].

Standard WSOL Technique

CAM(Class activation Mapping)[Zhou-2016]



Convolutional STN

Standard convolution sampling



(a) Standard convolution with a fixed receptive field.









STN Module and its action[Jaderberg]

Joint Probability and Multi-scale Localization

Replacing the GAP and following layers with CSTN followed by a 1x1 conv(#channels = #classes) layer gives a joint probability over location and classes.

a. Basic network

b. Joint probability

Applying CSTN and 1x1 conv over multiple layers of a feature pyramid gives a joint pdf over scale, location and classes. We used FPN.







c. Multiscale



d. CSTN

Overall System



Additional Regularization Terms



Reduce Localization Bias

$$L_{scale}(x) = \max\left(0, \max_{l} p(s=s_1, l, c=c^*|x) - \max_{l} (p(s=s_2, l, c=c^*|x))\right)$$



Convolutional STN Localization - Qualitative Results



CSTN Localization - CUB-200-2011 Test Set

Method	Top-1 Loc	GT-Known Loc	Top-1 Classif
CAM[Zhou-2016]	41.00	71.13	-
HAS[Singh-2017]	44.67	73.32	76.64
ACoL[Zhang-2018]	45.92	75.30	71.90
SPG[Zhang-2018]	46.64	74.11	-
ADL[Choe-2019]	62.29	78.62	80.34
CSTN	49.04	76.06	78.46
CSTN Top-5(GT-Known)	-	81.14	-

CSTN Localization - ILSVRC Validation Set

Method	Top-1 Loc	GT-Known Loc	Top-1 Classif
CAM[Zhou-2016]	42.80	61.10	66.60
HAS[Singh-2017]	45.21	63.12	70.70
ACoL[Zhang-2018]	45.83	62.73	67.50
SPG[Zhang-2018]	48.60	64.24	-
ADL[Choe-2019]	48.43	63.72	75.85
CSTN	42.38	60.48	69.48
CSTN Top-5(GT-Known)	-	63.45	-

Conclusion

- We proposed Convolutional STN, a WSOL method with an explicit component for learning Localization.
- Additional regularizations to prevent the localization from shrinking to discriminative areas and Introduce competition among different levels of the pyramid.
- Validated the proposed system on the benchmark datasets CUB-200-2011 and ImageNet Localization.

Further improvements:

- Incorporating more layers in the pyramid so that small objects can be accurately localized; also improve the competition among levels.
- Different strategies for selecting the localization box apart from the max-scoring one; Top-5 Localization shows we have better localization boxes learned by convolutional STN.

Questions?

