



Deep Neural Networks In Fully Connected CRF For Image Labeling With Social Network Metadata



Chengjiang Long



Roddy Collins



Eran Swears



Anthony Hoogs

Jan 10th, 2019.

Motivation

Gumpert Apollo S



According to Wikipedia this is the 9th best accelerating production car in the world, behind the Ariel Atom.

Ground truth: car, indoor, structure, transport.

On Sunday Afternoon



My wife and I took the dog out to run around on the highway.

Ground truth: animal, dogs, female, people, structures

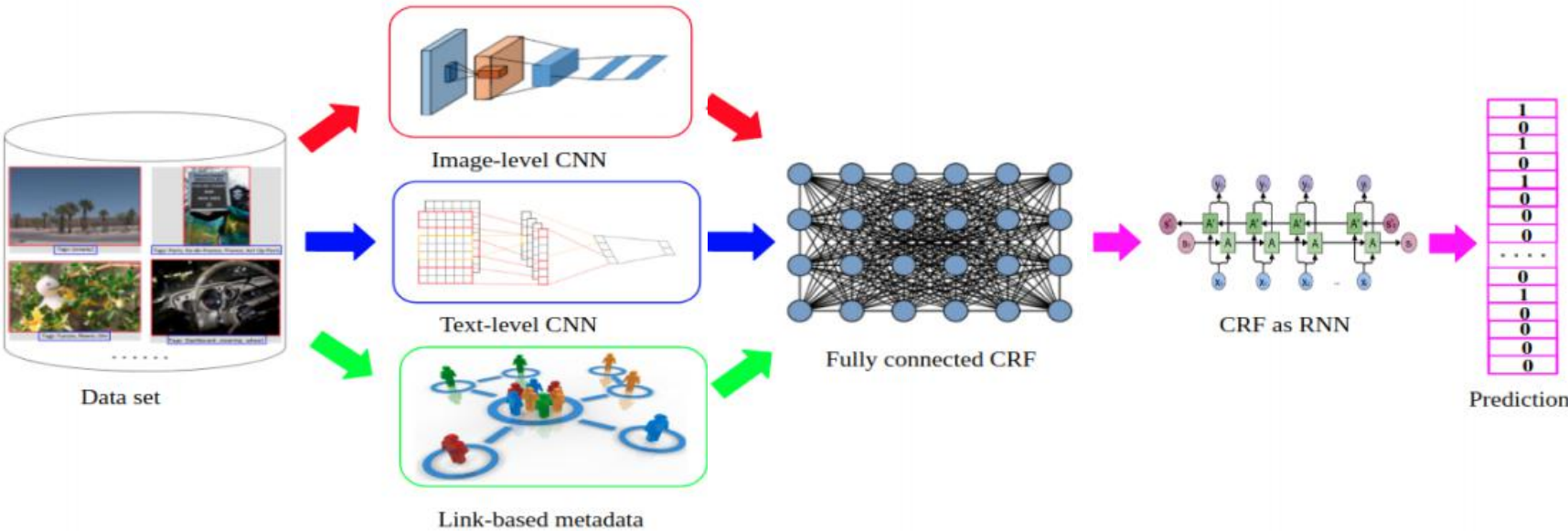
Observation: Social multimedia dataset contains

- (1) images
- (2) text information like title, description, comments
- (3) other meta information like user information, image gallery, uploader-defined groups, and links between shared contents.

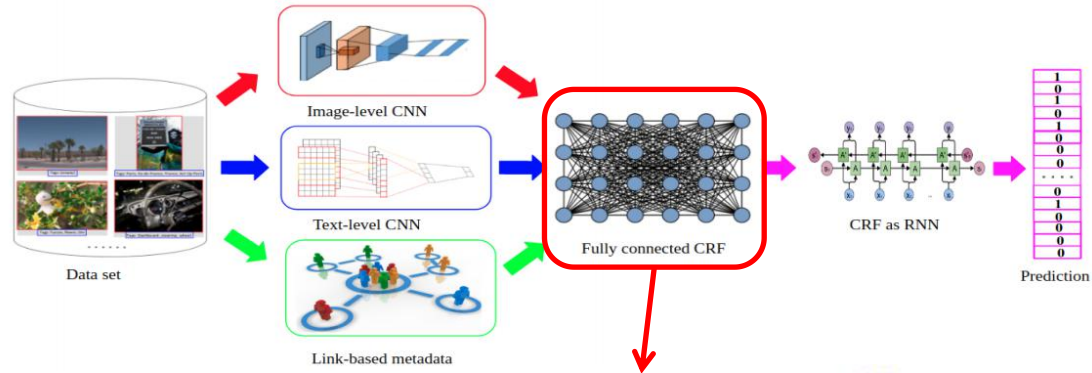
Intuition: We hypothesize that using social media context jointly with pixel information should improve the state-of-the-art in image labeling.

Goal: We seek to understand the relative contribution of pixels, text and other information in predicting image labels.

Proposed framework



CRF Formulation

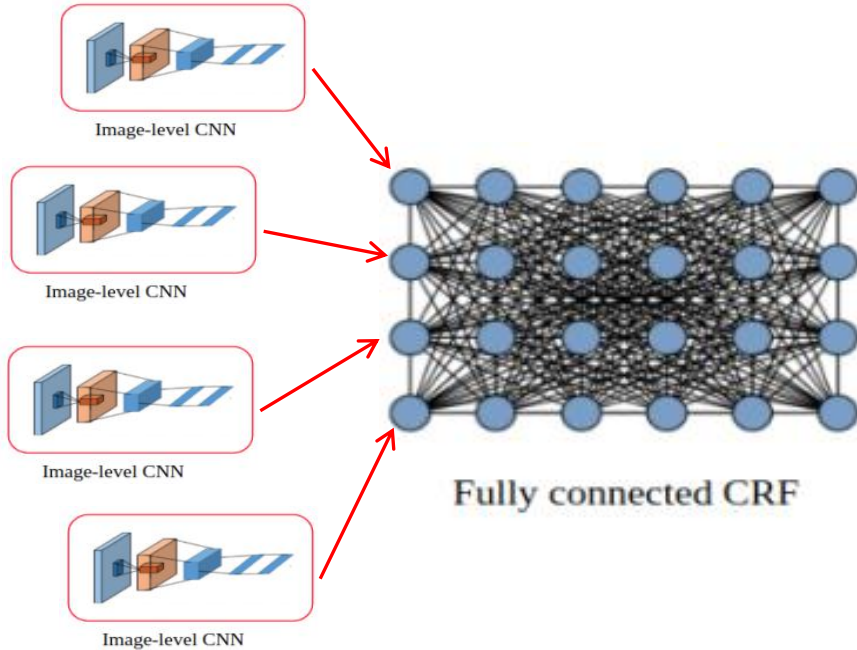


$$P(Y|I, M) = P(Y|\mathbf{x}, M) = \frac{1}{Z} \exp\left(\sum_{i=1}^N A(y_i, \mathbf{x}_i)\right) + \sum_{i=1}^N \sum_{\forall j \neq i} B(y_i, y_j, M)$$

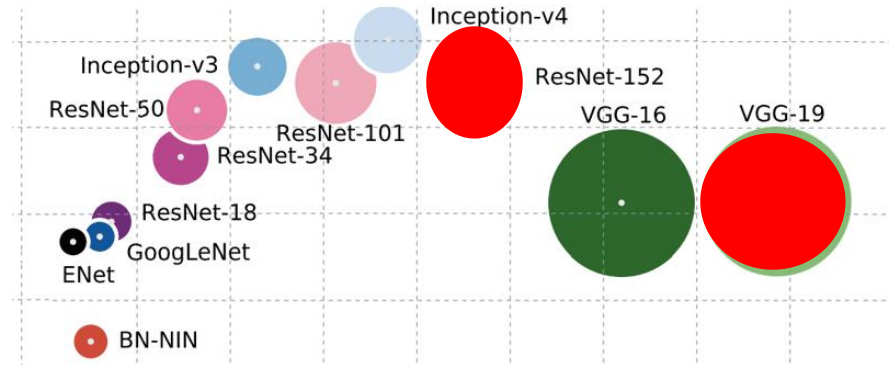
Unary potential

Pairwise potential

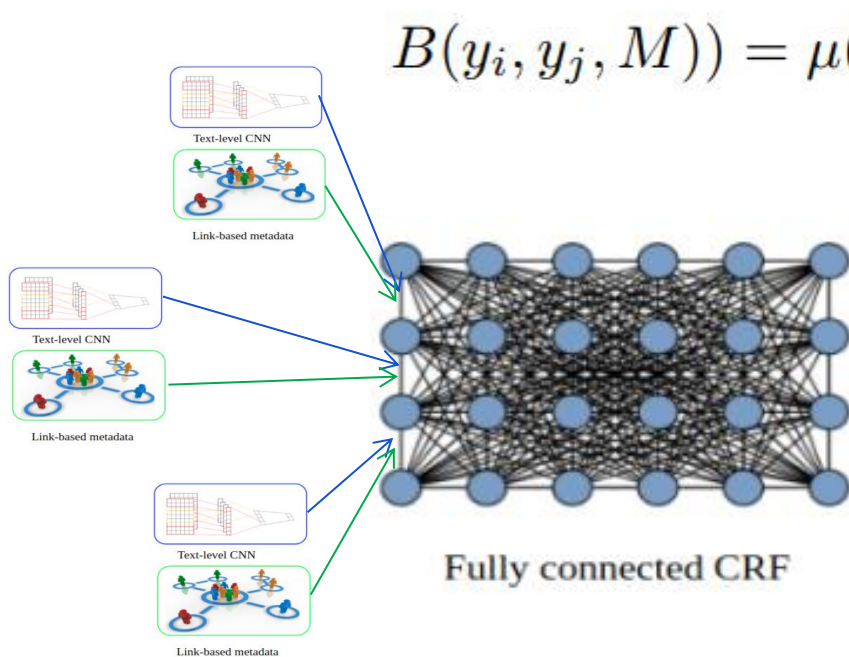
Unary potential with image-level CNN



$$A(y_i, \mathbf{x}_i) = \mathbf{w}_A^{y_i} \mathbf{x}_i + \mathbf{b}_A^{y_i}$$



Pairwise potential with text-level CNN and Metadata



$$B(y_i, y_j, M) = \mu(y_i, y_j) \sum_{k \in \{text, set, group\}} \mathbf{w}_B^k S_k(i, j)$$

$$S_{text}(i, j) = \exp\left(-\frac{|\mathbf{x}_i^{text} - \mathbf{x}_j^{text}|^2}{2\theta_{text}}\right)$$

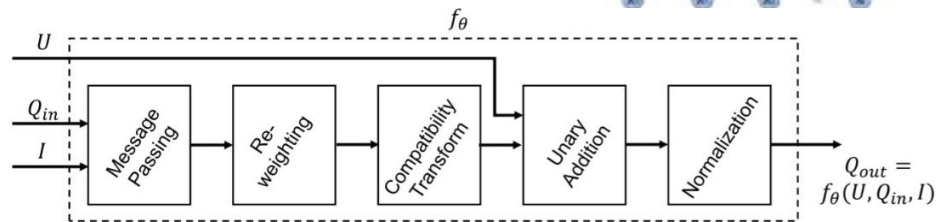
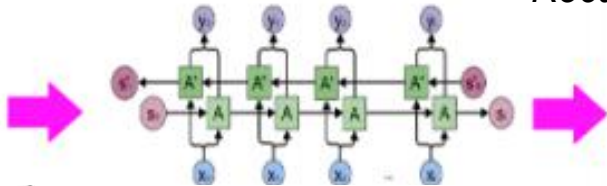
$$S_{set}(i, j) = \exp\left(-\frac{d(i, j, M_{set})^2}{2\theta_{set}}\right)$$

$$S_{group}(i, j) = \exp\left(-\frac{d(i, j, M_{group})^2}{2\theta_{group}}\right)$$

are

Training and inference

[S. Zheng et al. Conditional Random Fields as Recurrent Neural Networks. In ICCV 2015.]



$$Loss(Q, Y) = \sum_{i=1}^N -\frac{y_i}{N_+} \log Q_i(y_i) - \frac{1-y_i}{N_-} \log Q_i(1-y_i)$$

Binary cross entropy

$$+ \sum_{i=1}^N \frac{y_i}{N_+} (1 - (Q_i(y_i) - Q_i(1-y_i)))$$

$$+ \sum_{i=1}^N \frac{1-y_i}{N_-} (1 - (Q_i(1-y_i) - Q_i(y_i))) + \lambda \|w_A\|_2,$$

Ranking loss

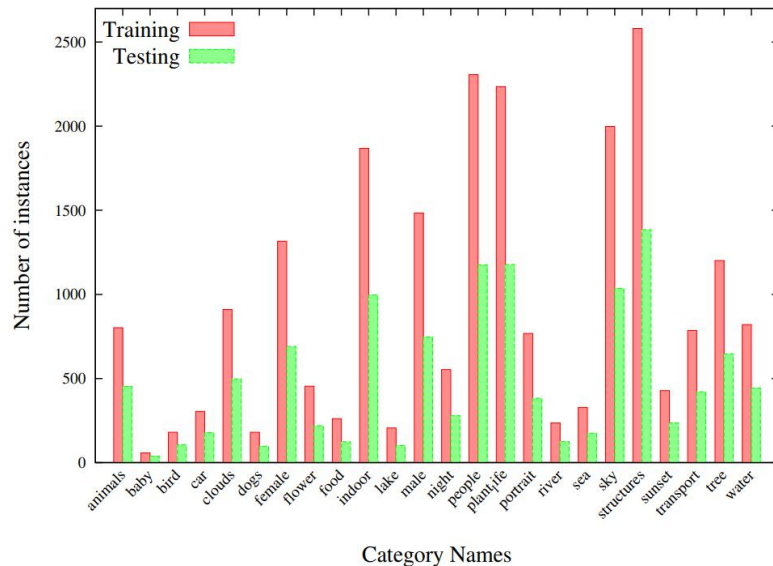
Algorithm 1: The outline of our proposed DCRF algorithm

Input: I and M
Output: Q

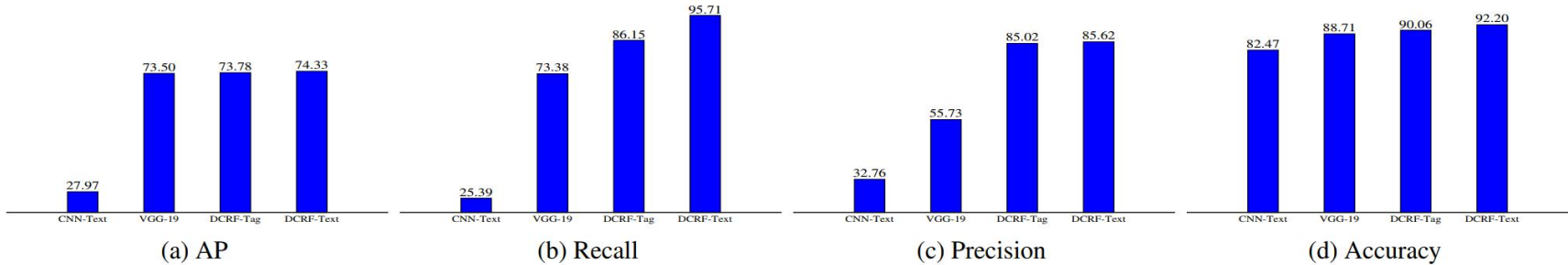
- 1 $x \leftarrow \text{CNN}_{image}(I)$
- 2 $x^{text} \leftarrow \text{CNN}_{text}(M)$
- 3 $t^{set}, t^{group} \leftarrow M$
- 4 $U \leftarrow w_A x + b_A$
- 5 $Q_i(y) \leftarrow \frac{1}{Z_i} \exp \{U_i(y)\}$
- 6 **while not converged do**
- 7 $\tilde{Q}_i^{(k)}(y) \leftarrow \sum_{\forall j \neq i} S_k(i, j) Q_j(y)$ for all k
- 8 $\check{Q}_i(y) \leftarrow \sum_k w_B^k \tilde{Q}_i^{(k)}(y)$
- 9 $\hat{Q}_i(y) \leftarrow \sum_{y'} \mu(y, y') \check{Q}_i(y)$
- 10 $\check{Q}_i(y) \leftarrow U_i(y) - \hat{Q}_i(y)$
- 11 $Q_i(y) \leftarrow \frac{1}{Z_i} \exp \{\check{Q}_i(y)\}$
- 12 **end**

MIR-9K Dataset

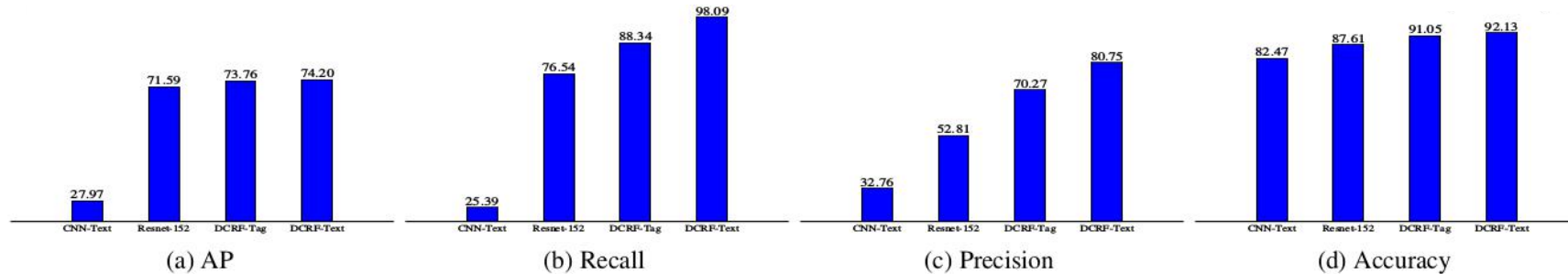
- A subset of the MIRFLICKR dataset.
- It contains 6000 training instances and 3182 testing instances with 24 categories.
- It involves a set of 3,213 users, a collection of 34,942 words and 17,687 image groups.



Effectiveness of text-level CNN

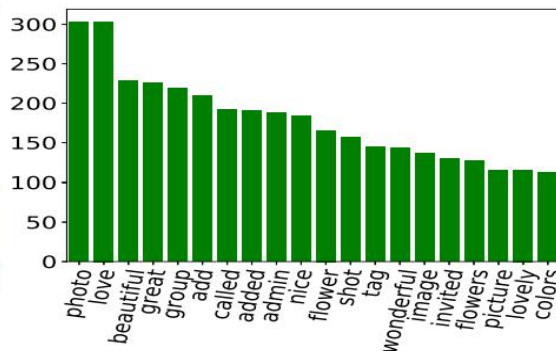
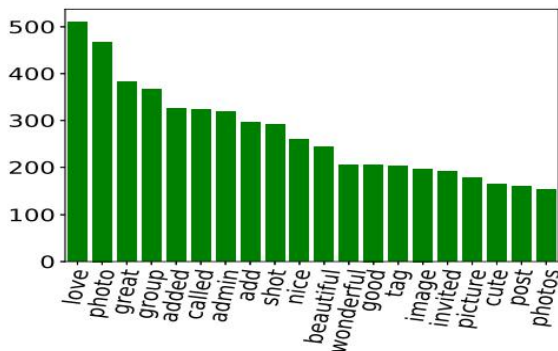
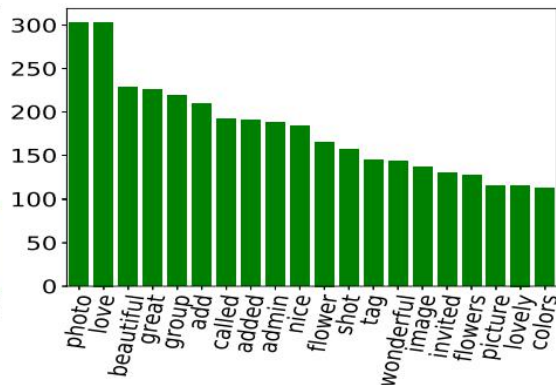
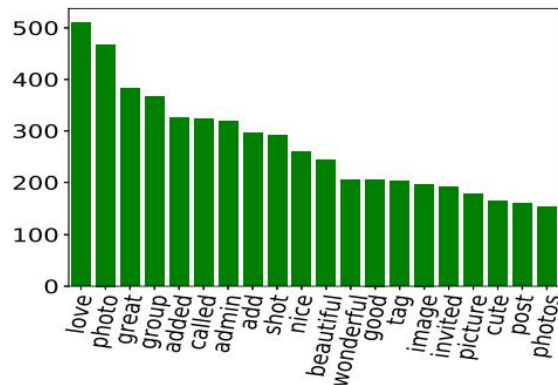


Node feature extracted from VGG-19 network.

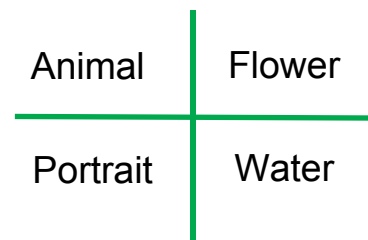


Node feature extracted from ResNet-152 network.

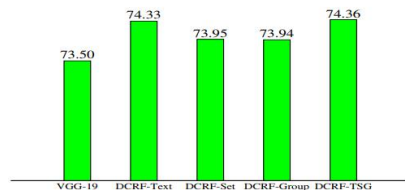
Visualization of top 20 tag words



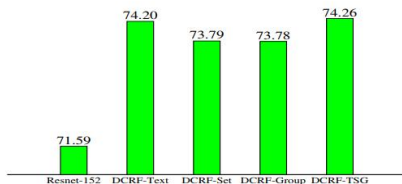
We observe that the top frequently cooccurring words such as “love”, “photo”, “great”, “group”, “added”, “nice” et. al convey little information relative to any of prediction of 24 categories.



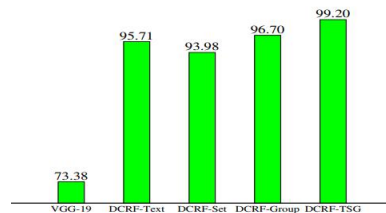
Effectiveness of the metadata for image labeling



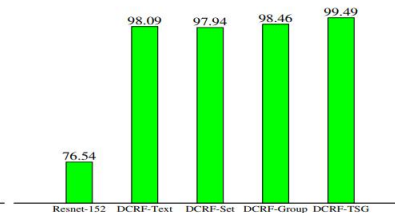
(a) AP with VGG-19



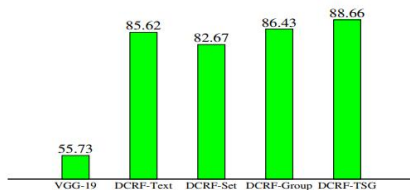
(b) AP with ResNet-152



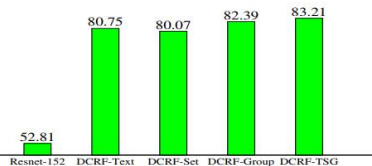
(c) Recall with VGG-19



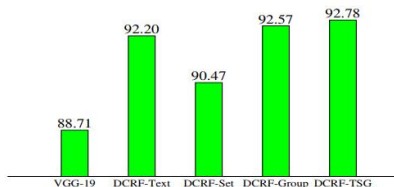
(d) Recall with ResNet-152



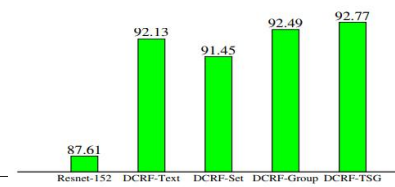
(e) Precision with VGG-19



(f) Precision with ResNet-152



(g) Accuracy with VGG-19



(h) Accuracy with ResNet-152

Compare with state-of-the-art approach

	AP	REC	PRE	ACC
CNN _{text} [15]	27.97	25.39	32.76	82.47
AlexNet _{img} [17]	62.54	76.30	40.25	74.56
VGG-19 _{img} [27]	73.50	77.38	55.73	88.71
ResNet-152 _{img} [9]	71.59	76.54	52.82	87.62
DenseNet-201 _{img} [10]	63.26	72.55	42.93	85.06
McAuley-CRF [21]	54.73	40.75	59.44	83.1
John-NCNN _{vgg} [12]	73.78	61.18	79.01	92.57
John-NCNN _{res} [12]	72.90	50.59	81.39	91.87
DCRF _{vgg} -BCE	74.13	92.66	85.86	92.50
DCRF _{vgg} -RLoss	74.29	93.12	88.18	92.61
DCRF _{vgg} -BCE+RLoss	74.36	99.20	88.66	92.78
DCRF _{res} -BCE	74.05	91.52	74.69	91.74
DCRF _{res} -RLoss	74.09	94.38	77.59	91.93
DCRF _{res} -BCE+RLoss	74.26	99.49	83.21	92.77

[12] J. Johnson et al. Love thy neighbors: Image annotation by exploiting image metadata. In ICCV 2015.

[15] Y. Kim. Convolutional neural networks for sentence classification. arXiv, 2014.

[21] J. J. McAuley and J. Leskovec. Image labeling on a network: Using social-network metadata for image classification. In ECCV, 2012.

Visualization

Taking in the Scenery



We took a walk through the woods and something too stressful since I was still a bit out of breath. I customized a few images for me. <https://www.flickr.com/photos/truhik/98/2057998788/> <http://www.bughugilabs.com/flickr/scout.php?mode=history&id=007-cjv-cjv> Woo hoo. Thanks guys!

VGG-19: animals, dogs, plant_life, structures
john-NCNN: animals, dogs, plant_life
DCRF: animals, dogs, plant_life
Ground-truth: animals, dogs, plant_life, tree

I think this is it



I looked through every night for something to set up and shoot. I've been having my cameras until I had INSPIRATION but it never came, so instead I forced it. I believe this to be the best shot I've taken. Get the lights on the way out.

VGG-19: female, food, indoor, male, people
john-NCNN: indoor
DCRF: indoor
Ground-truth: indoor

Sabo Bridge



On the bike-pedestrian bridge that crosses the downtown Greenway across Hiawatha Avenue and the Light Rail in Minneapolis. The bike is an Anon Fossil Fuel.

VGG-19: female, male, night, people, sky, structures, transport
john-NCNN: night, people, sky, structures, transport
DCRF: night, sky, structures, transport
Ground-truth: night, sky, structures, transport

White Orchid



White orchids seen in Nassau in a Gardens and Wildlife Center. A tiny number 20, 2007.

VGG-19: flower, lake, plant_life
john-NCNN: flower, plant_life
DCRF: flower, plant_life
Ground-truth: flower, plant_life

Sta of ik schiet-Hands Up



(Bijna) "De Haag - Die coron...
k...
z
j...
n...
acht onderdelen

VGG-19: car, female, flower, male, people, plant_life, portrait, structures
john-NCNN: female, male, people, structures
DCRF: male, people, structures
Ground-truth: male, people, structures

Homeless Sign



Tonight I went to a birthday/costume party and there was somebody dressed like a homeless guy. So I bummed his sign before I left.

VGG-19: female, indoor, male, people, portrait
john-NCNN: indoor, male, people
DCRF: indoor, male, people, portrait
Ground-truth: indoor, male, people, portrait

Alberto Granado

This very night... Granado...
born in Argentina in 1922. In 1952, with a broken bicycle, he made a long love across South America with the help of a doctor, Ernesto Guevara de la Serna, later nicknamed Che Guevara (but Granado still refers to him as 'es que'). They met at a diner in Havana, age 4 is a pleasure...
http://www.peepo.org/index.php?option=com_content&view=article&id=1629508...&Itemid=16
arrollo sig
nore (85 años de foto) è il dottore Alberto Granado nato in Argentina nel 1922. Nel 1952, con una moto rotta, si fece un lun go giro per l'America con un altro giovane (Ernesto Guevara de la Serna). Lui tardi soprannominato Che Guevara (ma Granado chiama solo il "marito").
a cena a casa e di stato un piacere discutere con...
http://it.wikipedia.org/index.php?title=Alberto_Granado&oldid=11472750...&Itemid=1629

VGG-19: female, indoor, male, people, portrait
john-NCNN: indoor, male, people
DCRF: indoor, male, people, portrait
Ground-truth: indoor, male, people, portrait

All done.



Subtitle: <...>
<...>
>Driven Car...
>ever...
>at seeme...
>like foreve...
>today. All of the...
>collaboration betwe...
>een myself, Jack...
>the dirty work). Serina Br...
>ekke. Steve...
>personal imager...
>provided by Sierra Korh...
Cassandra Leopold . Wow. Teamwork, huh?

VGG-19: indoor, male
john-NCNN: indoor
DCRF: indoor
Ground-truth: indoor

Goal keepers dream...



every... Goal keeper wish many hands at time of a penalty shoot out (i.a.dance shot)

VGG-19: female, indoor, male, night, people, sunset
john-NCNN: female, indoor, male, night, people, sunset
DCRF: indoor, male, night, people, sunset
Ground-truth: indoor, male, night, people

6pm



taken from PKiN (west side), Warsaw, Poland Explorer's best # 30

VGG-19: car, clouds, night, river, sky, structures, sunset, transport
john-NCNN: car, clouds, night, sky, structures, transport
DCRF: car, clouds, night, river, sky, structures, sunset, transport
Ground-truth: clouds, night, sky, structures, sunset, transport

Late Evening In The Black Valley



Killarney, Co. Kerry, Ireland.

VGG-19: car, clouds, plant_life, sky, structures, tree
john-NCNN: plant_life, sky, structures, tree
DCRF: plant_life, river, sky, structures, tree
Ground-truth: plant_life, river, sky, structures, water

Kapok tree branches, showing thorns



This is a branch of a kapok tree...
Cajon, California...
fruit, a more...
the flower...
The n...
I need to feel the thorns, to make...
izes available.

VGG-19: animals, clouds, flower, plant_life, sky, tree
john-NCNN: animals, plant_life, sky, tree
DCRF: plant_life, river, sky, tree
Ground-truth: plant_life, sky, tree

Conclusion and future work

- A novel deep fully connected CRF based framework DCRF with neural networks for image labeling using social network metadata.
- A joint end-to-end CNN-RNN formulation, which combines the strengths of both CNNs and RNNs.
- Our future work includes investigating more effective meta information, and improving the efficiency of the current DCRF framework to handle more complicated real-world application problems.

Welcome to our Poster 332.

Thanks!

Email: chengjiang.long@kitware.com

Web: <http://www.chengjianglong.com>

