

Learning Rich Features for Image Manipulation Detection

(Peng Zhou, Xintong Han, Vlad I. Morariu Larry S. Davis)



SAPIENZA
UNIVERSITÀ DI ROMA



Students:

Alberto Coluzzi
Caterina Borzillo
Federica Cocci

July 26, 2021

References:

- Official Paper:

[Learning Rich Features for Image Manipulation Detection \(thecvf.com\)](https://arxiv.org/abs/1512.01197)

- Faster R-CNN implementation:

[FurkanOM/tf-faster-rcnn: Tensorflow 2 Faster-RCNN implementation from scratch supporting to the batch processing with MobileNetV2 and VGG16 backbones \(github.com\)](https://github.com/FurkanOM/tf-faster-rcnn)

- COCO dataset:

[COCO - Common Objects in Context \(cocodataset.org\)](https://cocodataset.org/)

- Script to tamper the dataset:

[RGB-N/demo.py at master · pengzhou1108/RGB-N \(github.com\)](https://github.com/pengzhou1108/RGB-N)



Topic: Image Manipulation Detection

Are they **real** or **fake** images?



FAKE 



SAPIENZA
UNIVERSITÀ DI ROMA

2 goals:

1. Binary classification
2. Bounding box regression

1. Is it manipulated?



No.



Yes.

2. Where is it manipulated?



Bounding box found.



SAPIENZA
UNIVERSITÀ DI ROMA

Why is it innovative?

Fusion of two streams: RGB and Noise

Average Precision

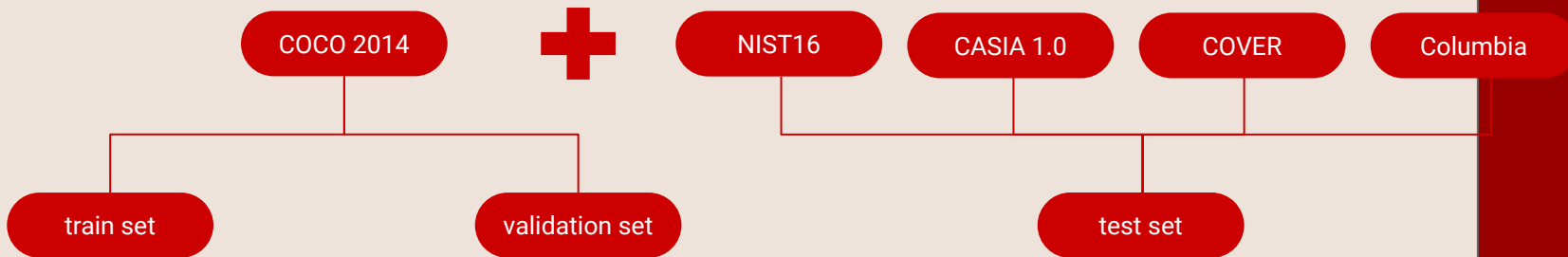
<u>AP</u>	Synthetic test
RGB Net	0.445
Noise Net	0.461
RGB-N noise RPN	0.472
Noise + RGB RPN	0.620
RGB-N	0.627

The two-stream network **RGB-N** performs **better** than each single stream!

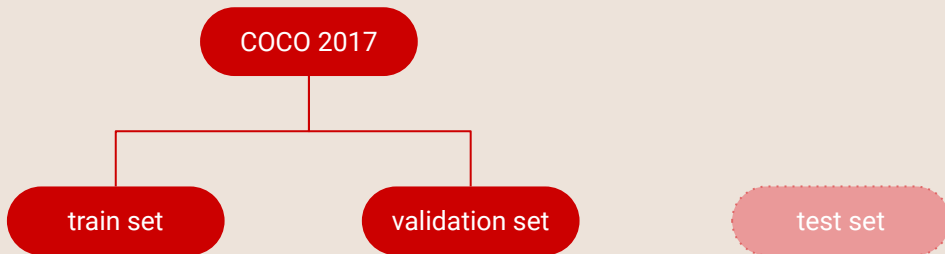
In addition, the **RPN** uses only **RGB features** that are more suitable with respect to noise features.



Dataset

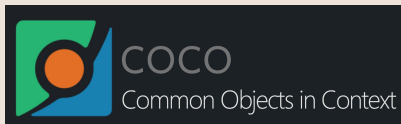


Paper implementation



Our implementation

Dataset: about COCO



What is COCO?



COCO is a large-scale object detection, segmentation, and captioning dataset.
COCO has several features:

- ✓ Object segmentation
- ✓ Recognition in context
- ✓ Superpixel stuff segmentation
- ✓ 330K images (>200K labeled)
- ✓ 1.5 million object instances
- ✓ 80 object categories
- ✓ 91 stuff categories
- ✓ 5 captions per image
- ✓ 250,000 people with keypoints

- MS COCO (Microsoft Common Objects in Context)
- The first version of MS COCO dataset was released in 2014.
- COCO 2017: training set composed by 118.000 images (18GB)



Dataset: manipulations



Obviously tampered

Difficult to say where it is tampered



Total of 78 classes and 26.415 images

```

class ImageManipulator:
    def __init__(self, image, mask, crop_box=None, scale_factor=1.0):
        self.image = image
        self.mask = mask
        self.crop_box = crop_box
        self.scale_factor = scale_factor

    def crop(self):
        if self.crop_box:
            x1, y1, x2, y2 = self.crop_box
            self.image = self.image[y1:y2, x1:x2]
            self.mask = self.mask[y1:y2, x1:x2]
        return self.image, self.mask

    def scale(self):
        if self.scale_factor != 1.0:
            new_size = (int(self.image.shape[0] * self.scale_factor),
                        int(self.image.shape[1] * self.scale_factor))
            self.image = cv2.resize(self.image, new_size)
            self.mask = cv2.resize(self.mask, new_size)
        return self.image, self.mask

    def apply(self):
        self.image, self.mask = self.crop()
        self.image, self.mask = self.scale()
        self.image = cv2.bitwise_and(self.image, self.image, mask=self.mask)
        return self.image
    
```

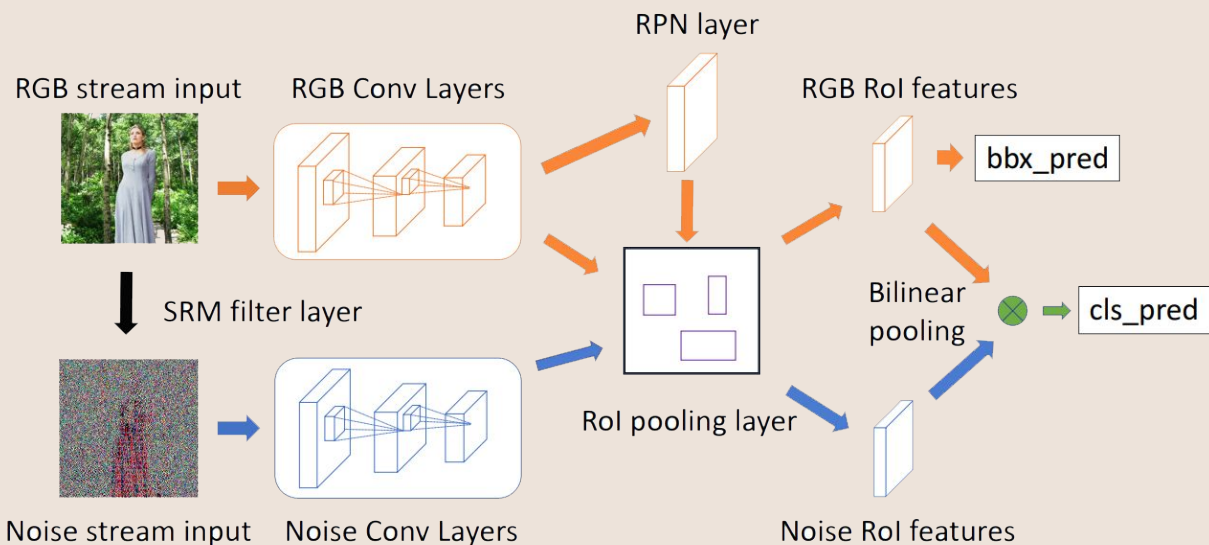
Script

Class	Number of Images
person	1279
snowboard	317
frisbee	322
bird	330
carrot	308
backpack	345
broccoli	319
tie	336
shis	352
site	315
spoon	344
chair	336
clock	356
apple	317
wine glass	319
elephant	379
keyboard	297
zebra	341
cake	338
dining table	335
airplane	368
fire hydrant	304
remote	302
horse	336
vase	320
sandwich	333
donut	307
orange	315
refrigerator	319
microwave	354
suitcase	313
dog	307
teddy bear	302
bottle	334
cell phone	315
book	331
bowl	324
car	338
baseball bat	326
oven	338
handbag	338
skateboard	318
plotted plant	322
bicycle	259
mousse	293
baseball glove	341
bear	370
laptop	306
fork	329
banana	315
surfboard	353
motorcycle	355
giraffe	337
train	373
bench	355
tennis racket	322
sports ball	332
tv	307
cat	298
truck	219
knife	335
stop sign	318
hair drier	324
bus	361
pizza	323
sink	338
scissors	304
cup	332
bed	305
cow	359
couch	312
sheep	364
toilet	333
hot dog	309
toaster	339
parking meter	306
umbrella	341
toothbrush	298



SAPIENZA
UNIVERSITÀ DI ROMA

Two-stream Faster R-CNN network



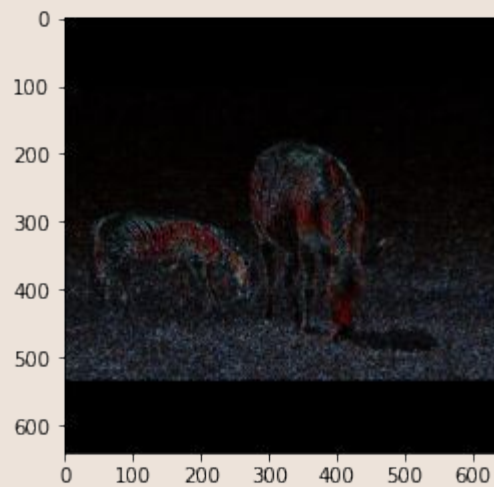
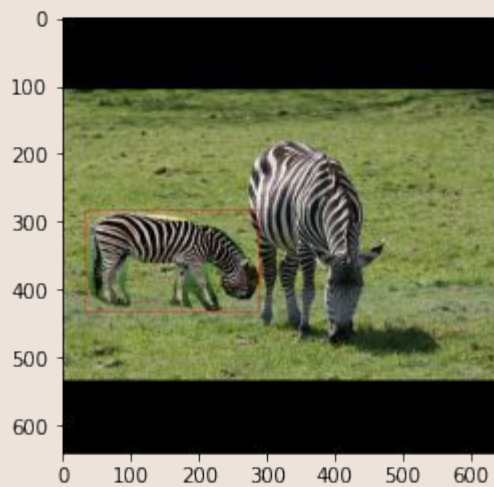
RGB stream:

- models **visual tampering artifacts**
- regresses **bounding boxes** to the ground-truth

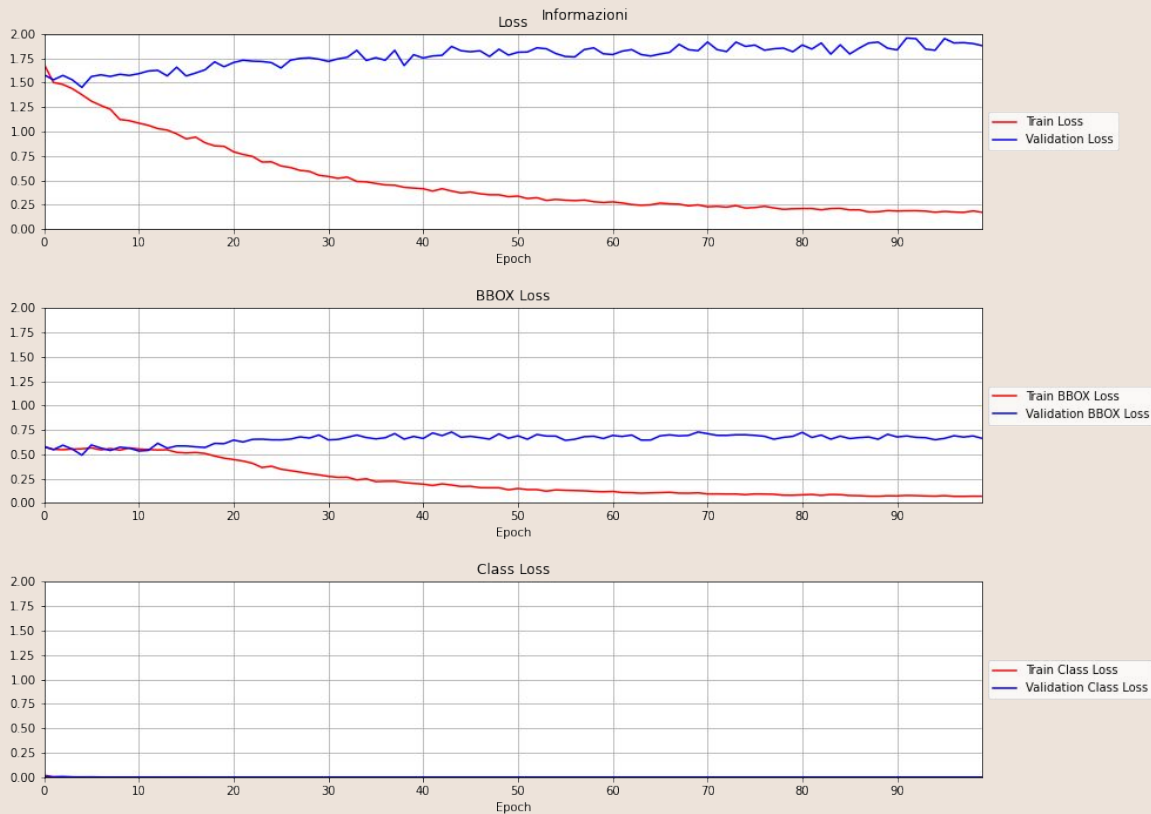
Noise stream:

- applies an **SRM filter** to the RGB image
- uses the noise features to provide **additional evidence** for manipulation classification





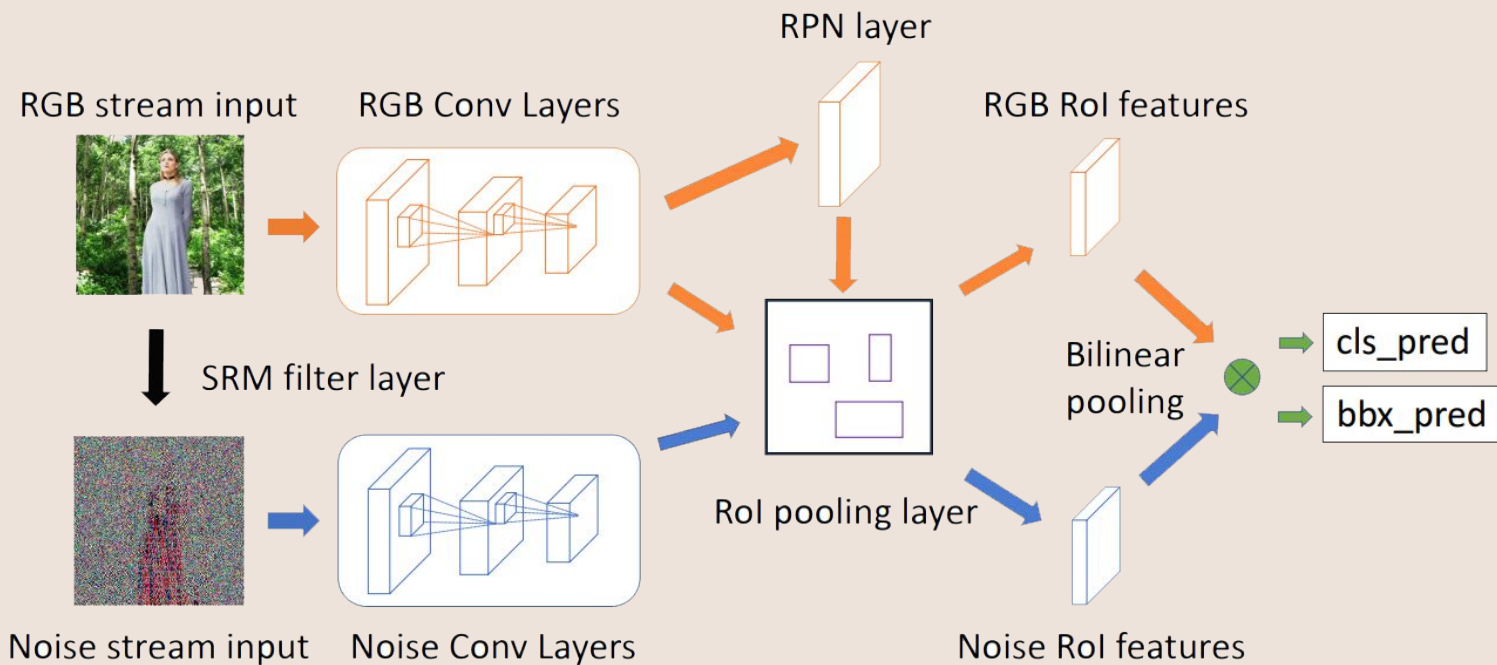
Results



Why this results ?



Our Model Flow



What we have learned ?

