Dataset	CASIA v2.0			Columbia			RT			Carvalho			FantasticReality		
	mAP	p-mAP	cIOU	mAP	p-mAP	cIOU	mAP	p-mAP	cIOU	mAP	p-mAP	cIOU	mAP	p-mAP	cIOU
ManTra	0.40	0.40	0.45	0.48	0.48	0.58	**0.50**	0.50	0.54	0.33	0.33	0.38	0.57	0.57	0.73
Ours	**0.74**	**0.74**	**0.76**	**0.69**	**0.69**	**0.77**	0.50	**0.51**	**0.55**	**0.48**	**0.48**	**0.56**	**0.61**	**0.61**	**0.76**

Table 1: **Splice Localization:** MAG vs. `ManTra-Net` [1] following the protocol defined in supplementary material.

Dear reviewers, we very much appreciate your valuable comments, time, and effort. Below we provide a detailed response to each reviewer.

R6: In terms of significant, the baselines the paper compares to for the forged region detection task, which is the main task considered in the paper, appear to be simple methods based on local statistics analysis.

We compare our MAG framework to two modern state-of-the-art deep learning splice detection frameworks: `LSC` is based on a Siamese network and was presented at ECCV 18, `MFCN` is a Multi-Task Fully Convolutional Network that was published in 2018 ([1,2] in our paper). We provide a comparison to non-deep learning methods to be consistent with the `LSC` paper. We compared our MAG framework to new deep learning-based `ManTra-Net` [1] that was not published during the submission and will be presented at CVPR 19 (see Table 1). Our method outperforms three state-of-the-art deep learning frameworks in terms of mAP, permuted mAP, and per class Intersection over Union (cIOU).

R6: Vague descriptions are abundant in the manuscript. The notation is also hard to follow. For example, Equation (2) and (3) are difficult to understand. It is unclear how the label loss is computed.

We are sorry for typos in equation (2) and (3), there should be '-', not ',' as it is noted by Reviewer 4. Equation (2) and (3) define a classical L1 distance. We will appreciate specific comments on the vague descriptions in the paper. We will clarify them in the camera-ready version. The class label loss is given by equation (6). The specific predicted class labels are given by $\hat{\boldsymbol{C}}_i$ for $i \in \{3, 4, \ldots, 2 + K\}$, where K is the number of classes ($K = 10$ in our experiments).

R5 / R6 / R4: The paper could benefit from more analysis on why it is so effective. I think the image-to-image translation part does not add much to the paper.

We can remove the image-to-image translation part and provide an extended analysis of the of (1) failure cases and (2) why is our framework is so effective. Examples in Figures 1 and 2 demonstrates how our retoucher G_R gradually removes the tampering artifacts in the input splice A with an increasing epoch. While other deep learning splice detection methods receive both realistic and rough splices from the first training epoch, our annotator G_A sees only rough splices at the first epoch. With an increasing epoch, retoucher G_R produces more complicated splices, which allows G_A to focus attention on the sophisticated tampering techniques that could appear in real splices. We believe that this is the main reason why our MAG framework achieves the state-of-the-art results and outperforms other deep learning methods.

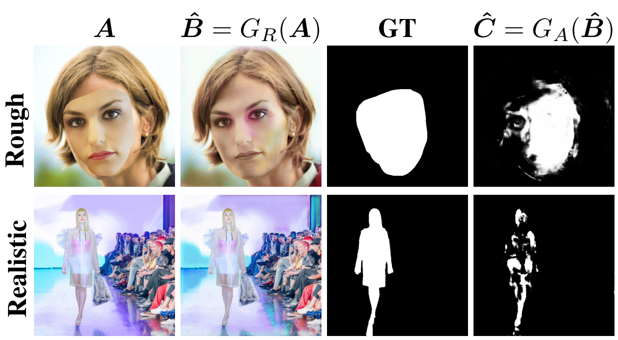

Figure 1: Performance for retoucher G_R on rough and realistic splices.

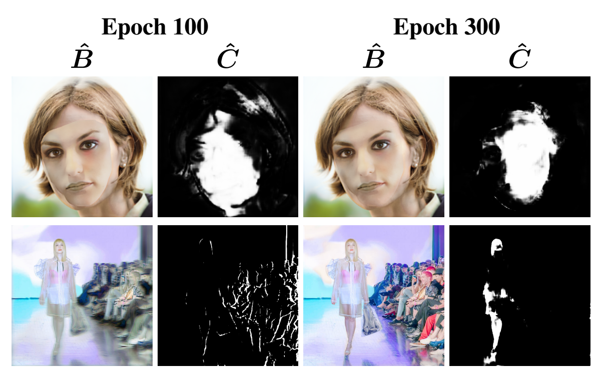

Figure 2: Adaptation of annotator G_A over time.

References

[1] Wael AbdAlmageed Yue Wu and Premkumar Natarajan. Mantra-net: Manipulation tracing network for detection and localization of image forgerieswith anomalous features. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.