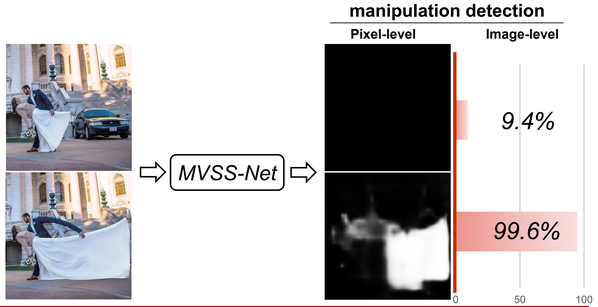
Image Manipulation Detection by Multi-View Multi-Scale Supervision

Xinru Chen* , Chengbo Dong* , Jiaqi Ji, Juan Cao, Xirong Li
AIMC Lab, Renmin University of China - Institute of Computing Technology, CAS

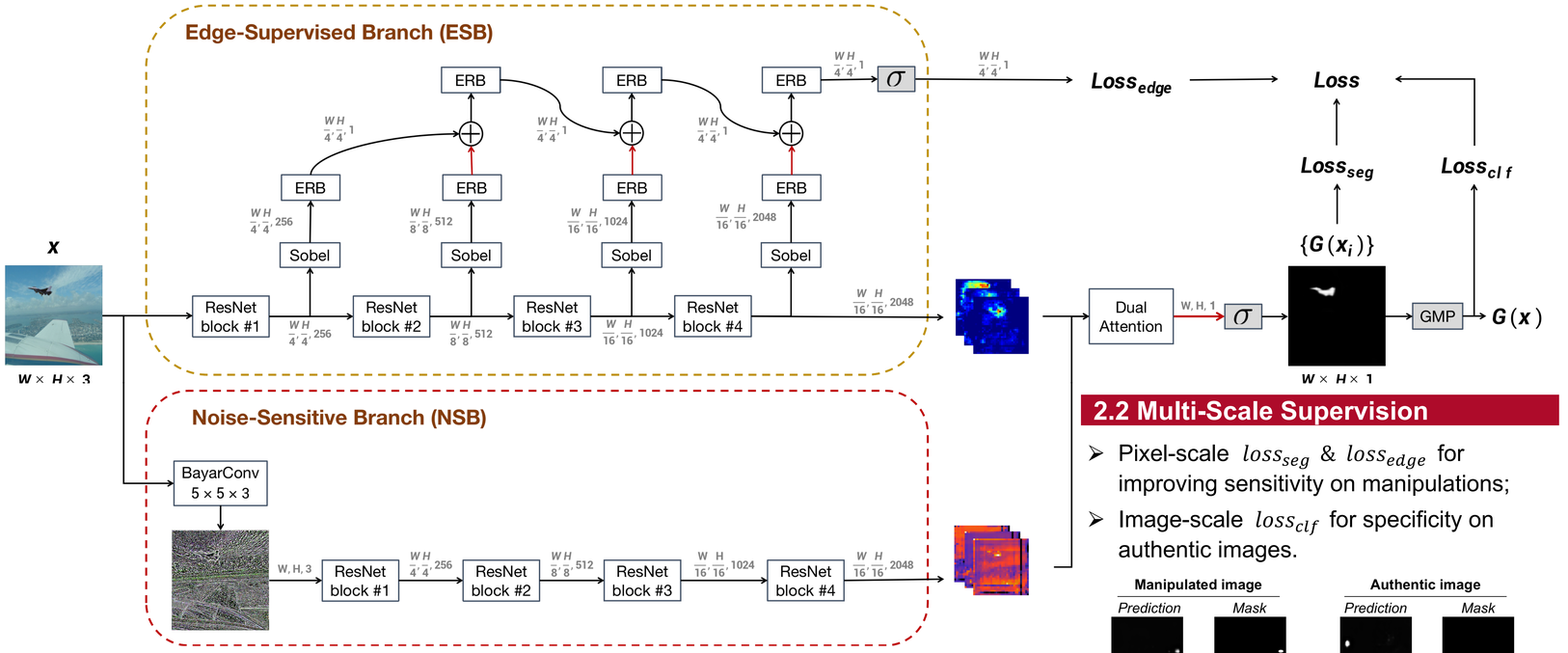
2021 ICCV OCTOBER 11-17 VIRTUAL

1. Summary

- **Background**: Digital images are easily manipulated in a misleading manner.
- **Challenges**
  1. Semantic-aware features lack **sensitivity** and generalizability in manipulated images;
  2. Authentic images are ignored by the prior art, leading to false alarm and poor **specificity**.
- **Our Solution**:
  1. Multi-view feature learning for sensitivity on manipulations;
  2. Multi-scale supervision to prevent false alarms on authentic.
- **Our novelty**: ***One model to obtain high sensitivity and specificity !***



2. Proposed MVSS-Net



2.1 Multi-View Feature Learning

- Edge and noise views contain manipulation traces;
- A shallow-to-deep edge supervision branch with Sobel enhancement produced a more focused response;
- Improve sensitivity on manipulations.

2.2 Multi-Scale Supervision

- Pixel-scale $loss_{seg}$ & $loss_{edge}$ for improving sensitivity on manipulations;
- Image-scale $loss_{clf}$ for specificity on authentic images.

	Manipulated image (Prediction / Mask)	Authentic image (Prediction / Mask)
Pixel-scale Loss	BCE Loss = 0.0198 DICE Loss = 0.2457 Effective for extremely imbalanced data	BCE Loss = 0.0286 Marginal contribution to reduce false alarms DICE Loss Inapplicable by definition
Image-scale Loss	BCE Loss(0.9647, 1) = 0.0359	BCE Loss(0.9961, 0) = 5.5413

Pixel-scale loss is not suitable for the authentic !

3. Main Results

3.1 Ablation Study on DEFACTO

Fixed threshold (0.5)

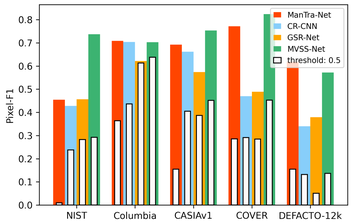
Setup	Pixel(F1) cpmv.	Pixel(F1) spli.	Pixel(F1) inpa.	Pixel(F1) MEAN	Image Sen.	Image Spe.	Image F1	Com-F1
Seg	0.45	0.72	0.46	0.55	0.83	0.62	0.71	0.62
Seg+Clf	0.34	0.67	0.38	0.46	0.77	0.78	0.77	0.58
Seg+Clf+Noise	0.39	0.71	0.43	0.51	0.76	0.82	0.79	0.62
Seg+Clf+Edge	0.41	0.72	0.44	0.52	0.77	0.81	0.79	0.63
Seg+Clf+GSRNet	0.36	0.71	0.42	0.50	0.81	0.78	0.80	0.61
Full setup	0.45	0.71	0.46	0.54	0.80	0.80	0.80	**0.64**

- Multi-scale supervision (Seg+clf) is less sensitive on manipulation;
- Both edge (Seg+Clf+Edge) and noise (Seg+Clf+Noise) are helpful;
- Proposed ESB is superior to previous feature concatenation (Seg+Clf+GSRNet).

3.2 Generalization study on public benchmarks against SOTA

Sensitivity on manipulation

- Using fixed threshold instead of optimal one calculated according to groundtruth is strict yet more practical.

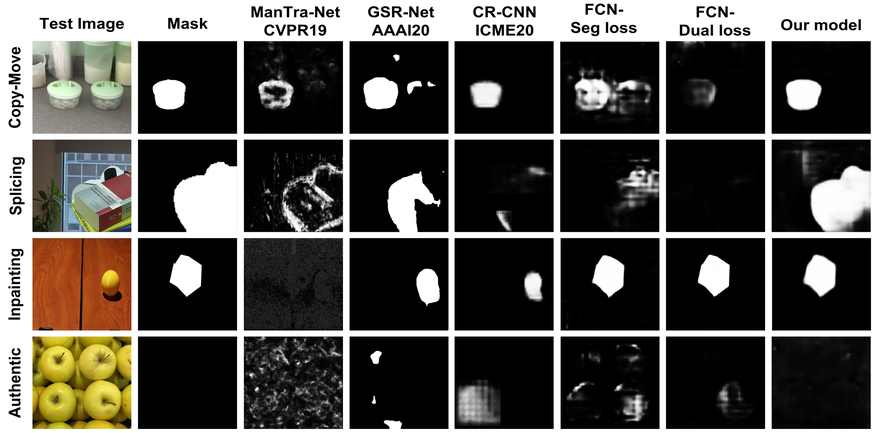


Specificity on authentic

Method	Columbia Sen.	Columbia Spe.	Columbia F1	CASIAv1 Sen.	CASIAv1 Spe.	CASIAv1 F1	COVER Sen.	COVER Spe.	COVER F1	DEFACTO-12k Sen.	DEFACTO-12k Spe.	DEFACTO-12k F1
ManTra-Net	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00
CR-CNN	0.96	0.25	0.39	0.93	0.22	0.36	0.97	0.07	0.13	0.77	0.27	0.40
GSR-Net	1.00	0.01	0.02	0.99	0.01	0.02	1.00	0.00	0.00	0.91	0.00	0.00
MVSS-Net	0.67	1.00	**0.80**	0.62	0.97	**0.75**	0.94	0.14	**0.24**	0.82	0.27	**0.41**

- SOTA works have poor specificity caused by false alarms. A joint evaluation on both Image-level and pixel-level F1 represents models' performances under real scenario.

Visualization



Github page: https://github.com/dong03/MVSS-Net