The Forchheim Image Database for Camera Identification in the Wild

Benjamin Hadwiger, Christian Riess

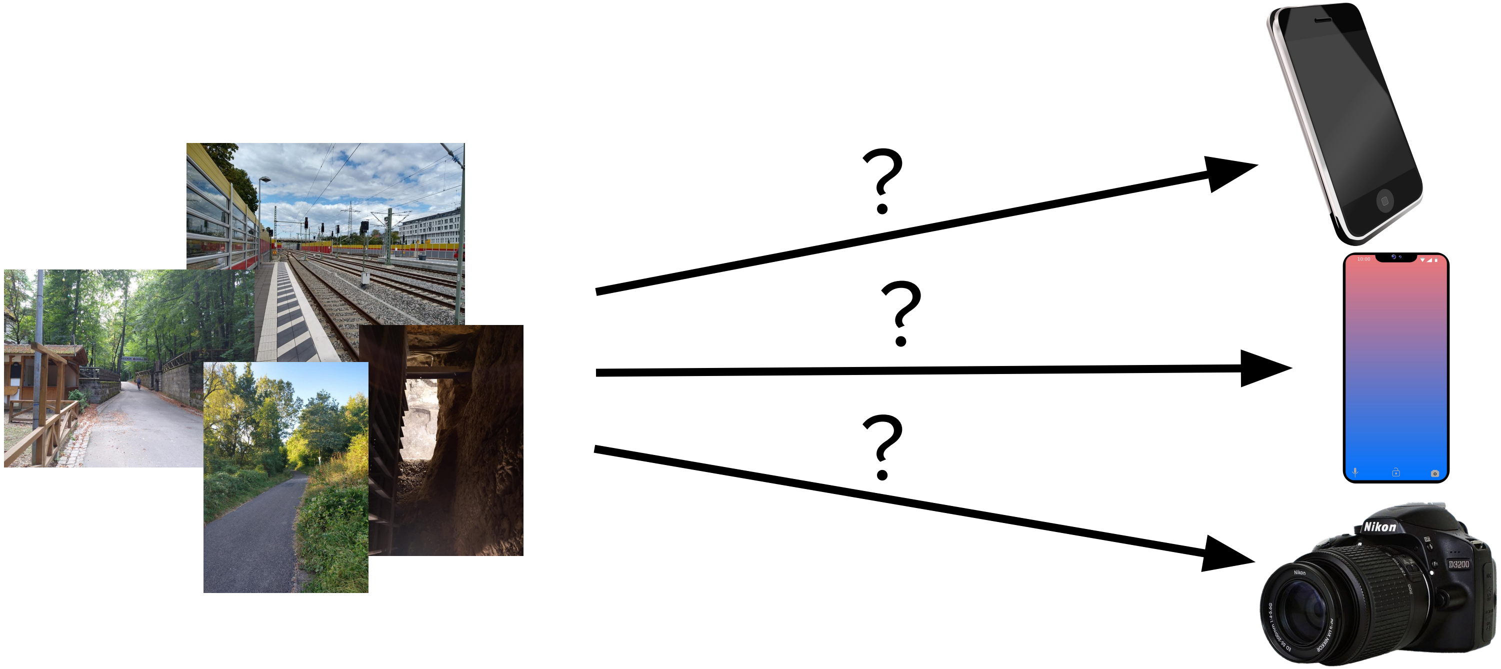
{benjamin.hadwiger, christian.riess}@fau.de

Multimedia Security Group, School of Engineering
Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

Which Camera Recorded the Images?

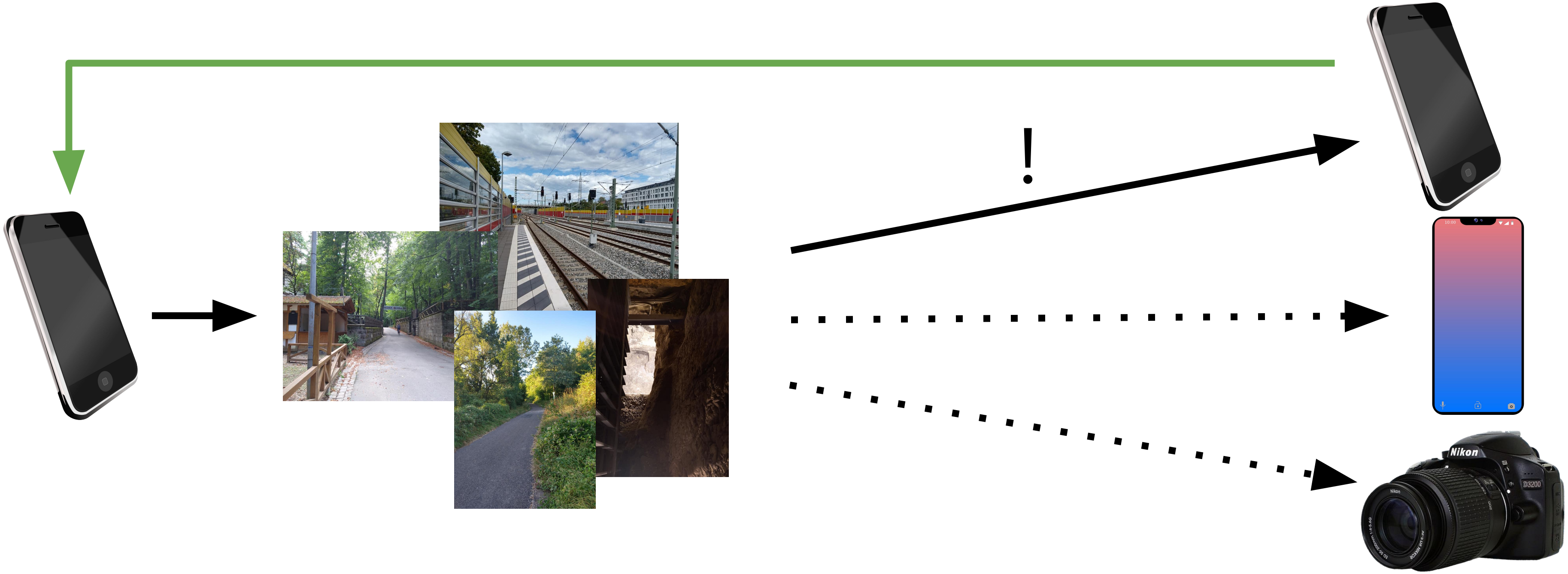
Which Camera Recorded the Images?

Knowledge on source camera can be important for criminal investigation



Which Camera Recorded the Images?

Camera can be accurately identified with existing methods



Which Camera Recorded the Images?

What about images shared online?

Which Camera Recorded the Images?

Most existing methods fail on strongly compressed images

How to Evaluate an Algorithm's Performance?

How to Evaluate an Algorithm’s Performance?

Scene split important for rigorous camera identification evaluation [1,2], *(our paper: Sec. 5.4)*

[1] Kirchner, Gloe: “Forensic Camera Model Identification“ (Handbook Dig. Forensics, 2015)
[2] Bondi *et al.*: “First Steps Toward Camera Model Identification With CNNs (Sig. Proc. Letters, 2017)

How to Evaluate an Algorithm's Performance?

Scene split important for rigorous camera identification evaluation [1,2], *(our paper: Sec. 5.4)*

Dresden Image Database (DIDB) [3]:

≈ 17,000 images of
83 scenes recorded by
73 devices
1 quality (orig.)

[1] Kirchner, Gloe: "Forensic Camera Model Identification" (Handbook Dig. Forensics, 2015)
[2] Bondi *et al.*: "First Steps Toward Camera Model Identification With CNNs (Sig. Proc. Letters, 2017)
[3] Gloe, Böhme: "The Dresden Image Database for Benchmarking Digital Image Forensics (J. Dig. Forensic Pract., 2010)

How to Evaluate an Algorithm's Performance?

Scene split important for rigorous camera identification evaluation [1,2], *(our paper: Sec. 5.4)*

Robustness considerations important to assess real-world applicability

Dresden Image Database (DIDB) [3]:
≈ 17,000 images of
83 scenes recorded by
73 devices
1 quality (orig.)

[1] Kirchner, Gloe: "Forensic Camera Model Identification" (Handbook Dig. Forensics, 2015)
[2] Bondi *et al.*: "First Steps Toward Camera Model Identification With CNNs (Sig. Proc. Letters, 2017)
[3] Gloe, Böhme: "The Dresden Image Database for Benchmarking Digital Image Forensics (J. Dig. Forensic Pract., 2010)

How to Evaluate an Algorithm’s Performance?

Scene split important for rigorous camera identification evaluation [1,2], *(our paper: Sec. 5.4)*

Dresden Image Database (DIDB) [3]:
≈ 17,000 images of
83 scenes recorded by
73 devices
1 quality (orig.)

Robustness considerations important to assess real-world applicability

VISION Database [4]:
≈ 30,000 natural images of
arbitrary scenes recorded by
35 devices in
4 qualities (orig., FB x2, WA)

[1] Kirchner, Gloe: “Forensic Camera Model Identification“ (Handbook Dig. Forensics, 2015)
[2] Bondi *et al.*: “First Steps Toward Camera Model Identification With CNNs (Sig. Proc. Letters, 2017)
[3] Gloe, Böhme: “The Dresden Image Database for Benchmarking Digital Image Forensics (J. Dig. Forensic Pract., 2010)
[4] Shullani *et al.*: “VISION: A Video and Image Dataset for Source Identification” (J. Inf. Sec. 2017)

How to Evaluate an Algorithm's Performance?

Scene split important for rigorous camera identification evaluation [1,2], *(our paper: Sec. 5.4)*

Dresden Image Database (DIDB) [3]:

≃ 17,000 images of
83 scenes recorded by
73 devices
1 quality (orig.)

DIDB

supports splitting by scenes
does not support benchmarking robustness

Robustness considerations important to assess real-world applicability

VISION Database [4]:

≃ 30,000 natural images of
arbitrary scenes recorded by
35 devices in
4 qualities (orig., FB x2, WA)

VISION

does not support splitting by scenes
supports benchmarking robustness

How to Evaluate an Algorithm's Performance?

Scene split important for rigorous camera identification evaluation [1,2], *(our paper: Sec. 5.4)*

Robustness considerations important to assess real-world applicability

supports splitting by scenes

supports benchmarking robustness

Forchheim Image Database (FODB)

Database

The Forchheim Image Database

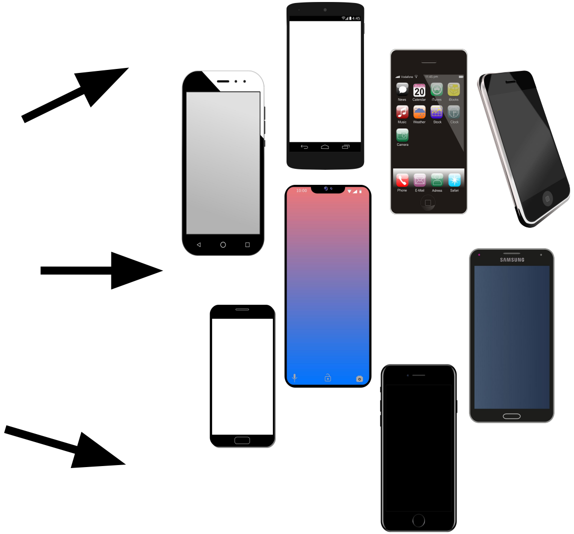
143 diverse scenes

Publicly available at: https://faui1-files.cs.fau.de/public/mmsec/datasets/fodb/

The Forchheim Image Database

143 diverse scenes

27 smartphones of
25 models, 9 brands



Publicly available at: https://faui1-files.cs.fau.de/public/mmsec/datasets/fodb/

The Forchheim Image Database (>23,000 Images)

143 diverse scenes

27 smartphones of
25 models, 9 brands

6 qualities
(orig. + 5 social networks)

Publicly available at: https://faui1-files.cs.fau.de/public/mmsec/datasets/fodb/

Experiments

Benchmarking Camera ID on FODB

We train the recent EfficientNet-B5 (EN-B5) [5],
a “general-purpose” CV CNN for camera ID

[5] Tan, Le: “Rethinking Model Scaling for CNNs” (ICML 2019)

Benchmarking Camera ID on FODB

We train the recent EfficientNet-B5 (EN-B5) [5],
a "general-purpose" CV CNN for camera ID

Forensic reference methods

- "BondiNet" (Bondi *et al.* [2])
- "MISLnet" (Mayer, Stamm [6])
- "RemNet" (Rafi *et al.* [7])

[2] Bondi *et al.*: "First Steps Toward Camera Model Identification With CNNs (Sig. Proc. Letters, 2017)
[5] Tan, Le: "Rethinking Model Scaling for CNNs" (ICML 2019)
[6] Mayer, Stamm: "Forensic Similarity for Digital Images" (TIFS 2020)
[7] Rafi *et al.*: "RemNet: Remnant CNNs for Camera Model Identification" (Neural. Comput. Appl. 2020)

Benchmarking Camera ID on FODB

We train the recent EfficientNet-B5 (EN-B5) [5], a "general-purpose" CV CNN for camera ID

Forensic reference methods

- "BondiNet" (Bondi *et al.* [2])
- "MISLnet" (Mayer, Stamm [6])
- "RemNet" (Rafi *et al.* [7])

Training / Evaluation Protocol

Data

- Scene split (train / val / test): 97 / 18 / 28
- 25 devices
- ➔ 2425 / 450 / 700 (train / val / test) images

[2] Bondi *et al.*: "First Steps Toward Camera Model Identification With CNNs (Sig. Proc. Letters, 2017)

[5] Tan, Le: "Rethinking Model Scaling for CNNs" (ICML 2019)

[6] Mayer, Stamm: "Forensic Similarity for Digital Images" (TIFS 2020)

[7] Rafi *et al.*: "RemNet: Remnant CNNs for Camera Model Identification" (Neural. Comput. Appl. 2020)

Benchmarking Camera ID on FODB

We train the recent EfficientNet-B5 (EN-B5) [5], a "general-purpose" CV CNN for camera ID

Forensic reference methods

- "BondiNet" (Bondi *et al.* [2])
- "MISLnet" (Mayer, Stamm [6])
- "RemNet" (Rafi *et al.* [7])

Training / Evaluation Protocol

Data

- Scene split (train / val / test): 97 / 18 / 28
- 25 devices
- ➔ 2425 / 450 / 700 (train / val / test) images
- Patch size
  - 64x64 (EN-B5, BondiNet, RemNet)
  - 256x256 (MISLnet)

[2] Bondi *et al.*: "First Steps Toward Camera Model Identification With CNNs (Sig. Proc. Letters, 2017)
[5] Tan, Le: "Rethinking Model Scaling for CNNs" (ICML 2019)
[6] Mayer, Stamm: "Forensic Similarity for Digital Images" (TIFS 2020)
[7] Rafi *et al.*: "RemNet: Remnant CNNs for Camera Model Identification" (Neural. Comput. Appl. 2020)

Benchmarking Camera ID on FODB

We train the recent EfficientNet-B5 (EN-B5) [5], a "general-purpose" CV CNN for camera ID

Forensic reference methods
- "BondiNet" (Bondi *et al.* [2])
- "MISLnet" (Mayer, Stamm [6])
- "RemNet" (Rafi *et al.* [7])

Results on "clean" image (accuracy [%])

CNN	64x64	256x256	Image
BondiNet	71.4	84.9	93.1
MISLnet	--	93.5	96.8
RemNet	93.8	96.6	**99.1**
EN-B5	**96.3**	**98.1**	**99.1**

[2] Bondi *et al.*: "First Steps Toward Camera Model Identification With CNNs (Sig. Proc. Letters, 2017)
[5] Tan, Le: "Rethinking Model Scaling for CNNs" (ICML 2019)
[6] Mayer, Stamm: "Forensic Similarity for Digital Images" (TIFS 2020)
[7] Rafi *et al.*: "RemNet: Remnant CNNs for Camera Model Identification" (Neural. Comput. Appl. 2020)

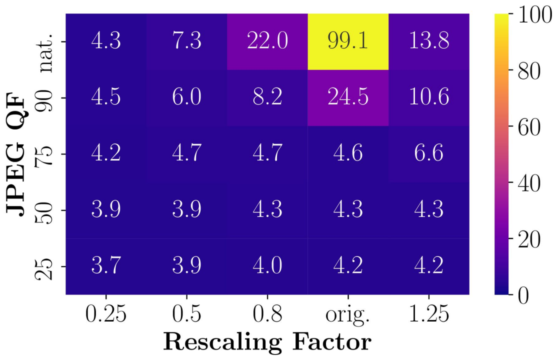
Stress-Testing CNN Robustness on FODB

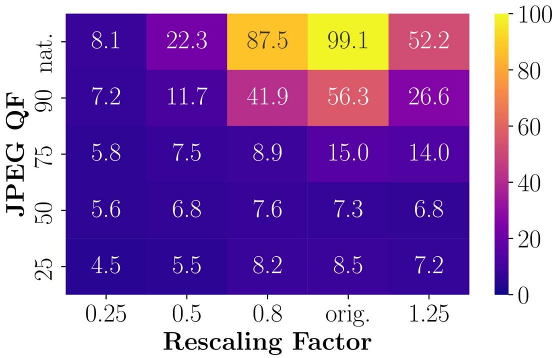
Evaluate for double JPEG compression and rescaling during test (image-level)

Stress-Testing CNN Robustness on FODB

Evaluate for double JPEG compression and rescaling during test (image-level)

RemNet





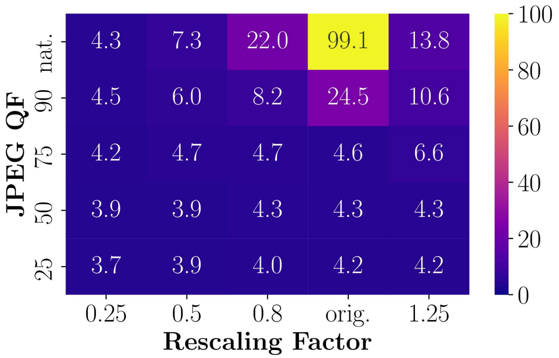
EN-B5

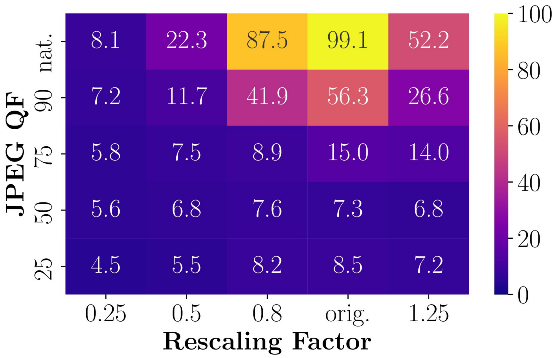
Stress-Testing CNN Robustness on FODB

Evaluate for double JPEG compression and rescaling during test (image-level)

Can we do better?

RemNet





EN-B5

Stress-Testing CNN Robustness on FODB

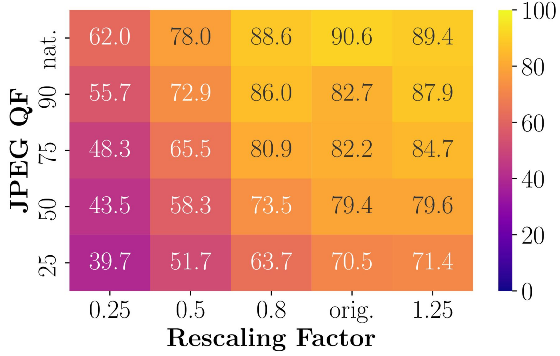
Training augmentation with strong degradations ("deg.")

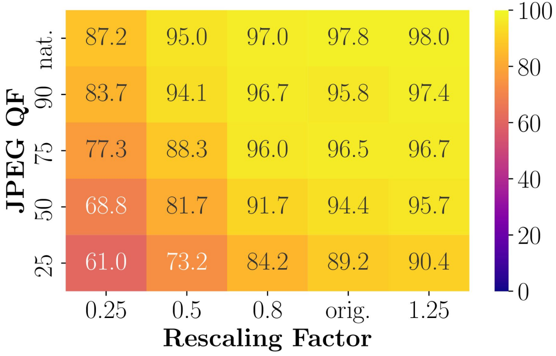
- Rescaling: $f \in [0.25, \ldots, 4.0]$
- JPEG recompression: $QF \in [100, \ldots, 10]$

OpenCV implementations

Stress-Testing CNN Robustness on FODB

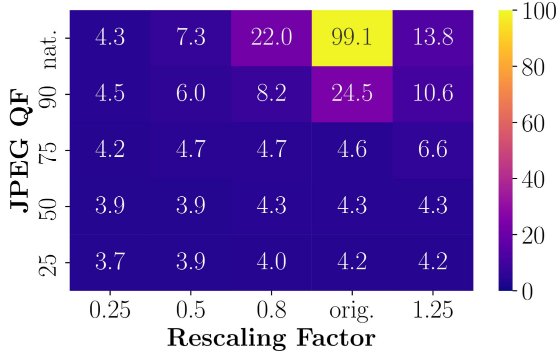
RemNet
with deg.

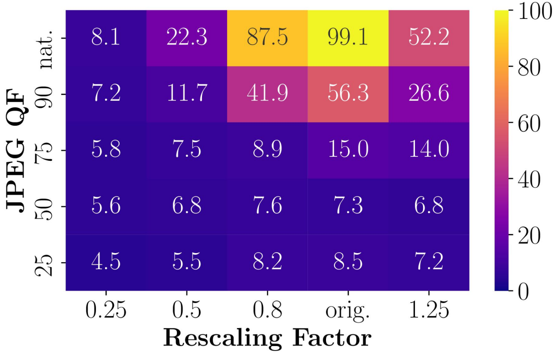




EN-B5
with deg.

RemNet
without deg.





EN-B5
without
deg.

Stress-Testing for Real-World Post-Processing on FODB

Evaluate on test images with unknown real-world post-processing (black box)

Social media versions of test images in FODB

➜ Facebook (FB), Instagram (IG), Telegram (TG), Twitter (TW), Whatsapp (WA)

Stress-Testing for Real-World Post-Processing on FODB

Evaluate on test images with unknown real-world post-processing (black box)

Social media versions of test images in FODB

→ Facebook (FB), Instagram (IG), Telegram (TG), Twitter (TW), Whatsapp (WA)

Comparison of 3 variants:

(1) Baseline: EfficientNet trained on orig. images without any degradation
(2) EfficientNet trained on orig. images with strong degradation (OpenCV impl.)
(3) Oracle: EfficientNet trained social media versions of *training split*

Stress-Testing for Real-World Post-Processing on FODB

(1) Baseline: EfficientNet trained on orig. images without any degradation
(2) EfficientNet trained on orig. images with strong degradation (OpenCV impl.)
(3) Oracle: EfficientNet trained on social media versions of *training split*

Training			**Test Dataset**					
	Dataset	**Degr.**	**orig**	**FB**	**IG**	**TG**	**TW**	**WA**
(1)	orig	no	99.1	4.6	5.6	5.3	9.8	6.8

Stress-Testing for Real-World Post-Processing on FODB

(1) Baseline: EfficientNet trained on orig. images without any degradation
(2) EfficientNet trained on orig. images with strong degradation (OpenCV impl.)
(3) Oracle: EfficientNet trained on social media versions of *training split*

Training			**Test Dataset**					
	Dataset	**Degr.**	**orig**	**FB**	**IG**	**TG**	**TW**	**WA**
(1)	orig	no	99.1	4.6	5.6	5.3	9.8	6.8
(2)	orig	yes	98.0	51.1	67.5	73.1	93.2	72.9

significant
improvement

Stress-Testing for Real-World Post-Processing on FODB

(1) Baseline: EfficientNet trained on orig. images without any degradation
(2) EfficientNet trained on orig. images with strong degradation (OpenCV impl.)
(3) Oracle: EfficientNet trained on social media versions of *training split*

Training			Test Dataset					
	Dataset	**Degr.**	**orig**	**FB**	**IG**	**TG**	**TW**	**WA**
(1)	orig	no	99.1	4.6	5.6	5.3	9.8	6.8
(2)	orig	yes	98.0	51.1	67.5	73.1	93.2	72.9
(3)	post-pr.	no	-	71.4	84.0	86.2	97.7	90.4

significant improvement

upper bound on performance

Stress-Testing for Real-World Post-Processing on FODB

(1) Baseline: EfficientNet trained on orig. images without any degradation
(2) EfficientNet trained on orig. images with strong degradation (OpenCV impl.)
(3) Oracle: EfficientNet trained on social media versions of *training split*

Training			**Test Dataset**					
	Dataset	**Degr.**	**orig**	**FB**	**IG**	**TG**	**TW**	**WA**
(1)	orig	no	99.1	4.6	5.6	5.3	9.8	6.8
(2)	orig	yes	98.0	51.1	67.5	73.1	93.2	72.9
(3)	post-pr.	no	-	71.4	84.0	86.2	97.7	90.4

significant improvement

upper bound on performance

Conclusion

We proposed FODB, a large-scale image forensics benchmark to jointly

- align training / test splits with scenes
- evaluate on real-world post-processing

Conclusion

We proposed FODB, a large-scale image forensics benchmark to jointly

- align training / test splits with scenes
- evaluate on real-world post-processing

We showed that artificial degradation can generalize to unknown post-processing

- FODB enables determining an upper bound on blind augmentation camera ID

Conclusion

We proposed FODB, a large-scale image forensics benchmark to jointly

- align training / test splits with scenes
- evaluate on real-world post-processing

We showed that artificial degradation can generalize to unknown post-processing

- FODB enables determining an upper bound on blind augmentation camera ID

FODB can also be used to evaluate social network provenance

- can be used to create a multi-image sharing dataset with aligned scenes

Conclusion

We proposed FODB, a large-scale image forensics benchmark to jointly

- align training / test splits with scenes
- evaluate on real-world post-processing

We showed that artificial degradation can generalize to unknown post-processing

- FODB enables determining an upper bound on blind augmentation camera ID

FODB can also be used to evaluate social network provenance

- can be used to create a multi-image sharing dataset with aligned scenes

Blind augmentation can also improve CNN-based forgery detection

Thank you!