Image Watermarking

Tomoki Tanimura (@tanimu)
d-hacks, Jin Nakazawa Lab, SFC, Keio University, M1

Created At: 2020/10/27

What is Watermark?

Protect Copyrights!

What is Watermark

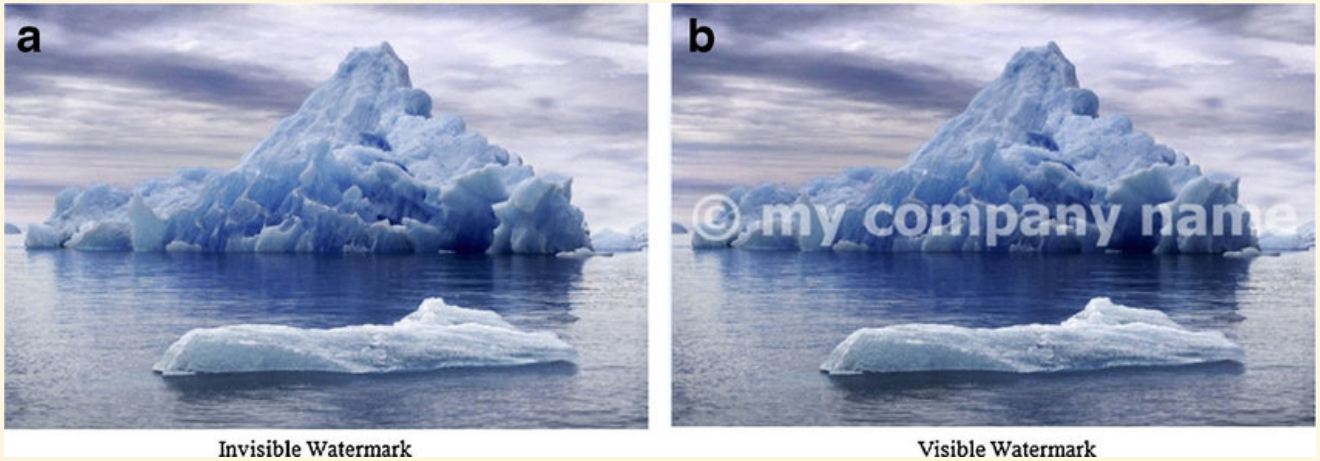
- Digital Watermark is the technology to leave the copyright information in an image itself

Steganography VS Watermark

- Watermark
  - protect copyright of the image
  - publisher leave a message to an image to protect the rights
- Steganograpy
  - send secret message to reciever
  - sender and receiver want to communicate without leaking to adversary

Visible and Invisible Watermark

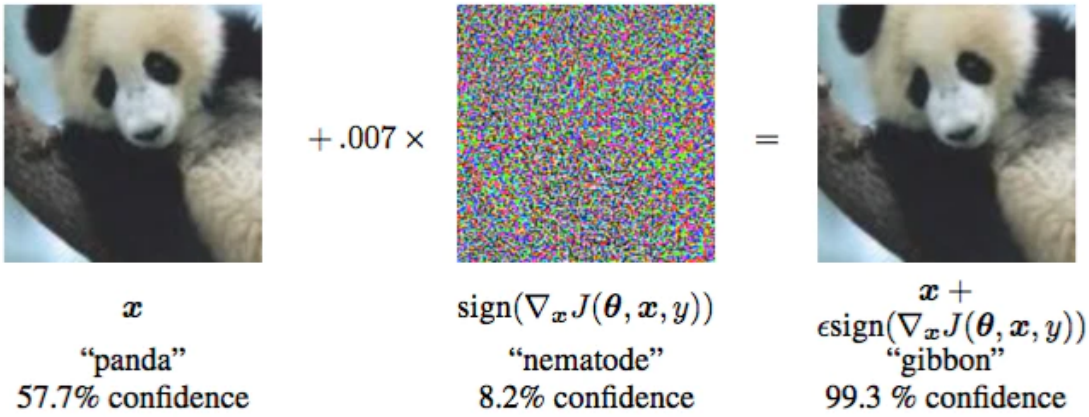
- Visible
  - Ordinary Watermark which often used
  - Easy to put it
- **Invisible**
  - Human cannot recognize watermark
  - Put watermark in the imperceptible space to human



Invisible Watermark Visible Watermark

Why Watermark with NN?

- NN is sensitive to tiny pertubation in image (GoodFellow, 2014)
  - This feature cause adversarial attack problem which cannot be solved yet
- We can use this sensitivity to hide and extract information in image



Invisible Watermark

You cannot see it!

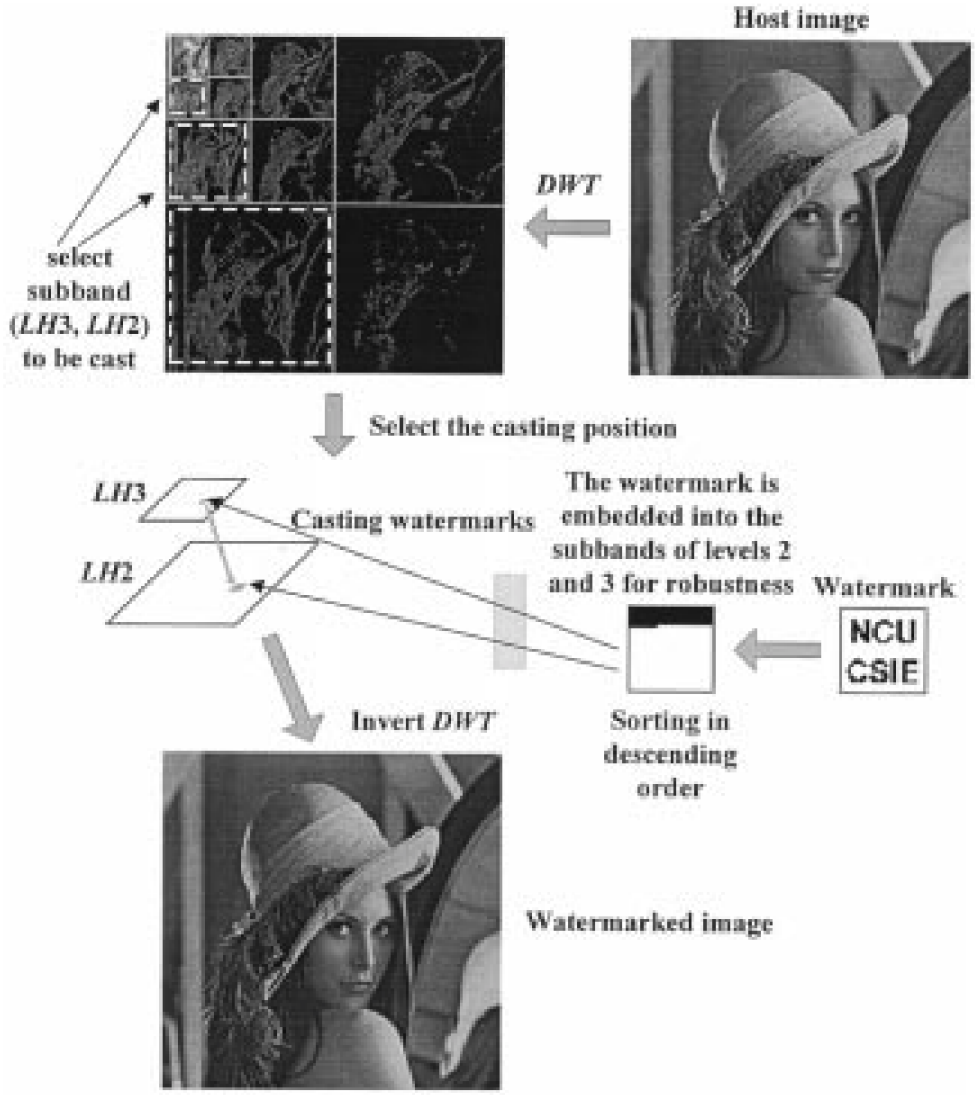
Traditional Methods

Fourier Transformation

- Robust to linear transformation (Shelby, 2000)

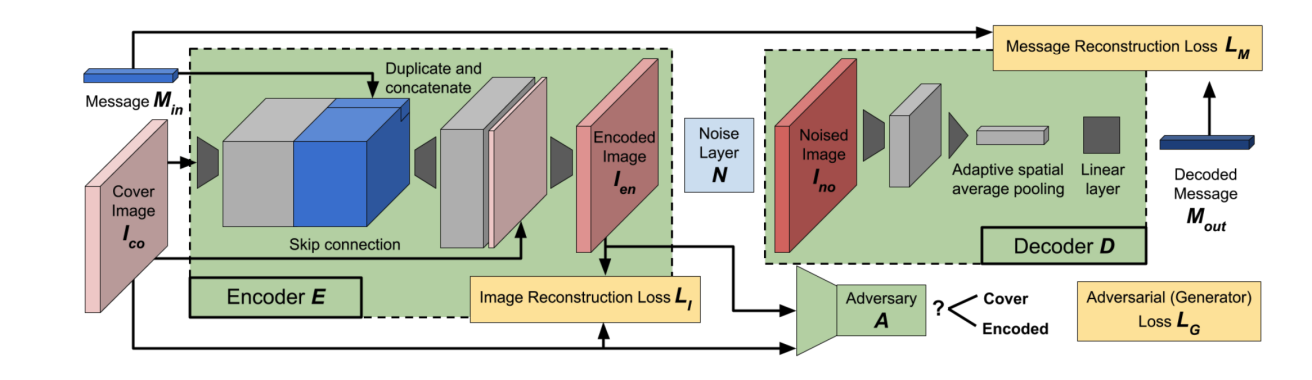
Wevelet Transformation

- Two-band wavelet transform, embed the watermark bits on the wavelet coefficients (Ning, 2007)
- Multiband one, embedd it in the mean trend of some middle-frequency subimages (Ming, 2001)

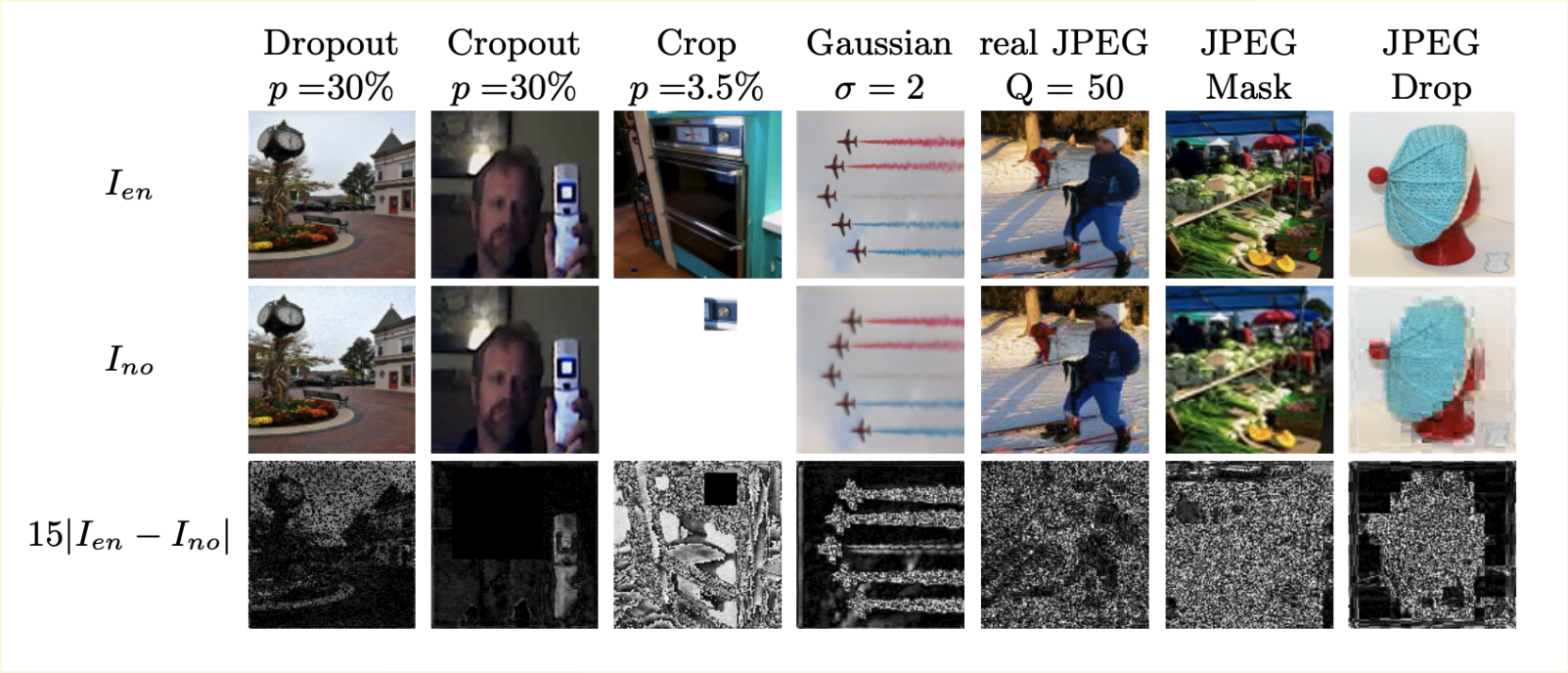

Host image
DWT
select
subband
(LH3, LH2)
to be cast
Select the casting position
LH3
LH2
Casting watermarks
The watermark is
embedded into the
subbands of levels 2
and 3 for robustness
Watermark
NCU
CSIE
Invert DWT
Sorting in
descending
order
Watermarked image

HiDDeN: Hiding Data with Deep Networks

- **More robust better image quality than traditionals**
- Model
  - **Encoder**: encode message to image
  - **Adversary**: Distort the encoded image
    - gaussian blur, dropout, crop, jpeg compression, ...
  - **Decoder**: Decode the encoded message from the image

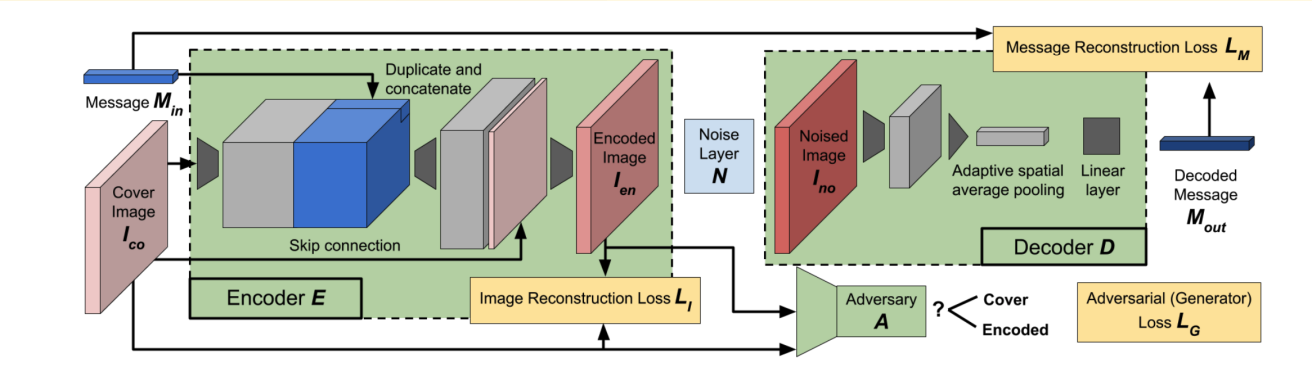


Adversary (Noise)



Loss Function

- 3 Loss fucntions
  - Image Reconstruction Loss
    - Difference between cover and encoded image
  - Message Reconstruction Loss
    - Difference between original and decoded message
  - Adversarial Loss
    - Adversarial Loss for cover and encoded image



Metrics

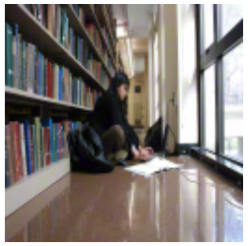
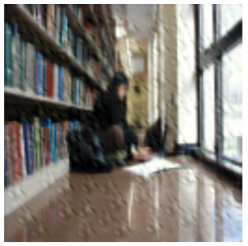
- Capacity
  - Bits per Pixel (`= L / HWC`)
- Secrecy
  - PSNR (larger is higher quality)
  - Detection Rate (for steganography)
- Robustness
  - Bit Accuracy of Message

Important metics for steganography and watermark are different
Robustness is most important for watermark

Experiment / Secrecy

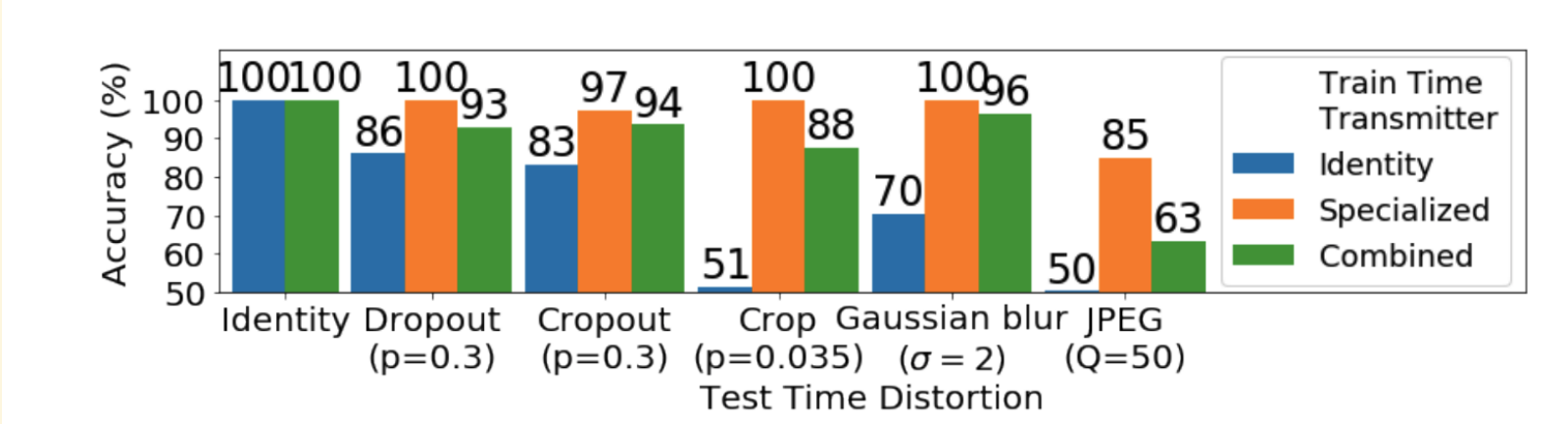
- Quality is not bad

	Digimarc	Identity	Dropout	Cropout	Crop	Gaussian	JPEG-mask	JPEG-drop	Combined
PSNR(Y)	62.12	44.63	42.52	47.24	35.20	40.55	30.09	28.79	33.55
PSNR(U)	38.23	45.44	38.52	40.97	33.31	41.96	35.33	32.51	38.92
PSNR(V)	52.06	46.90	41.05	41.88	35.86	42.88	36.27	33.42	39.38

		Trained with Adversary			No Adversary
Cover	Digimarc	Crop	Gaussian	Combined	Combined
					

Experiment / Robustness

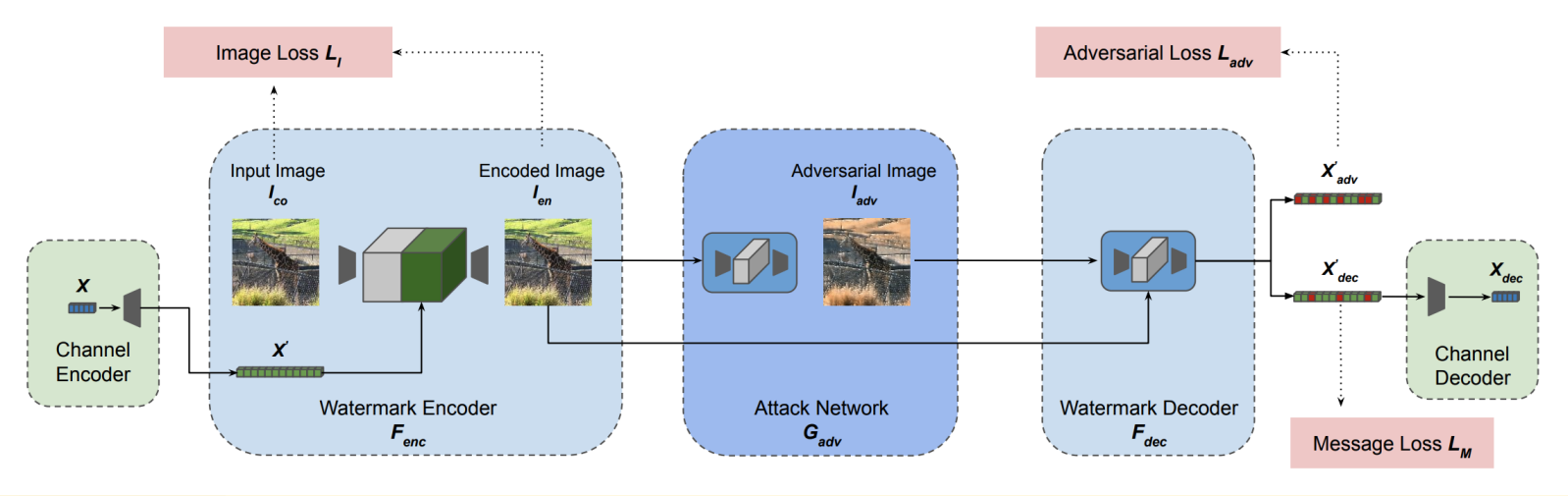
- JPEG is hardest distortion
- Specialized trained model is robust except for JPEG
- Non-adversary trained model cannot deal with Crop



Distortion Agnostic Deep Watermarking

CVPR2020

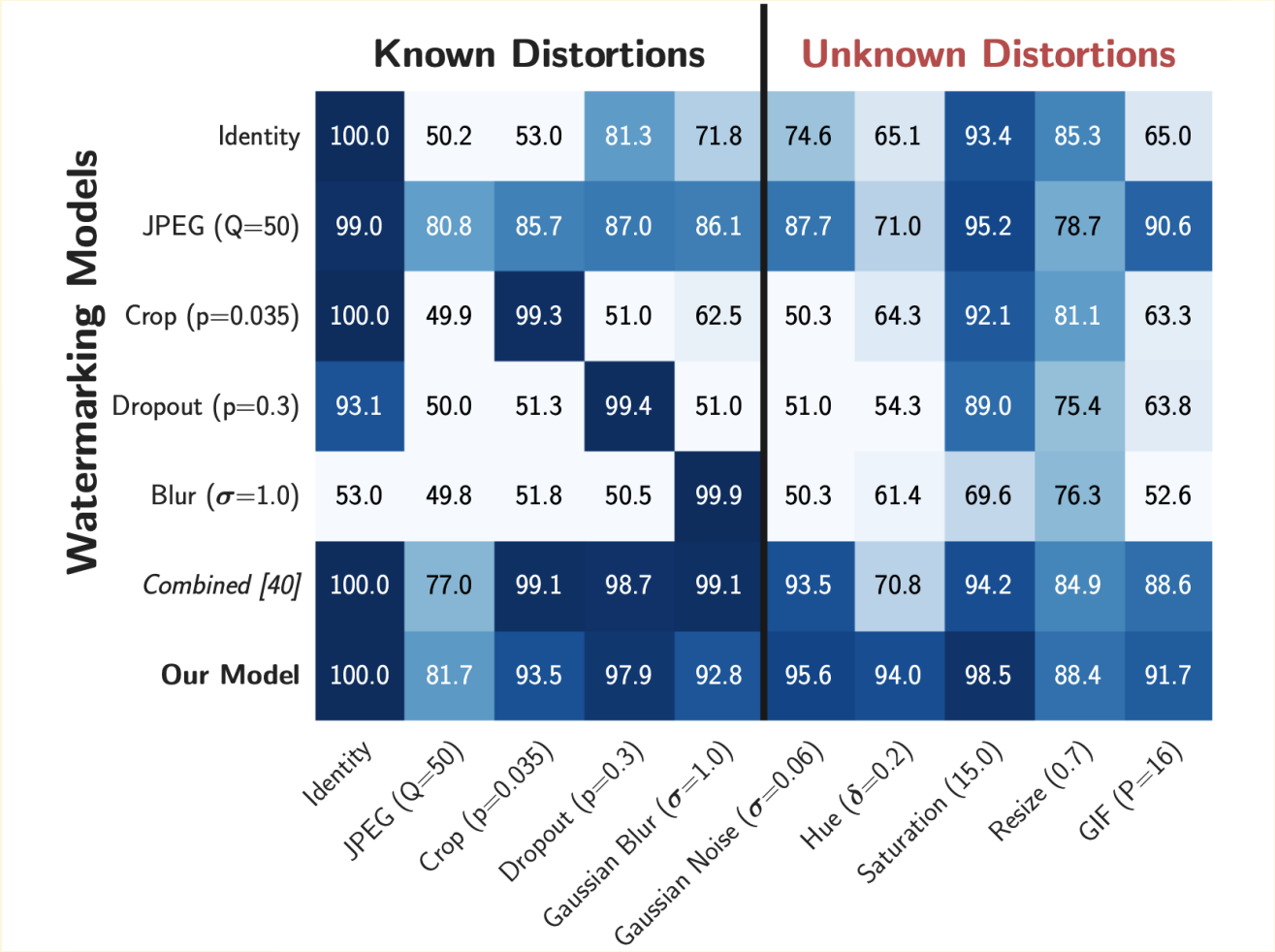
- "Attack Network" distort the encoded image with NN
  - Not define the specific distortion type
- "Channel Encoding" treat the message as the feature vector in the model



Result

Figure 2: Example of original image, encoded image and difference between the two images from our model.

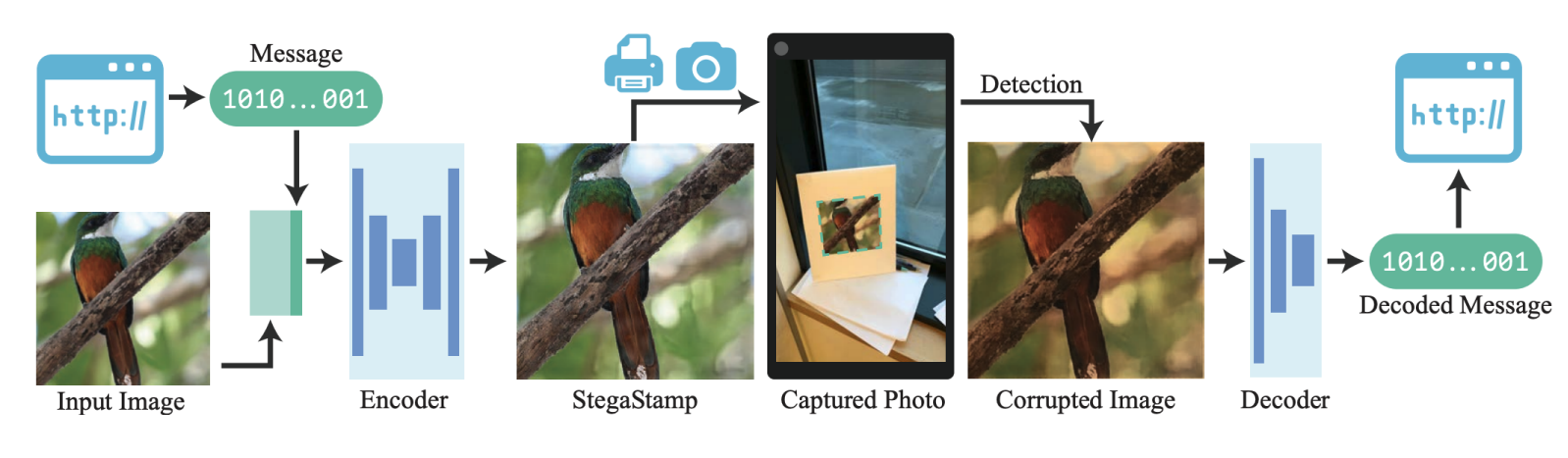
- As robust as the specialized and combined of HiDDeN



StegaStamp: Invisible Hyperlinks in Physical Photographs

CVPR2020

- Realize the tracebility of realworld phisycal images
- e.g. Embed URL to realworld images, and it can be scanned by AR glass



Encoded images

Realworld Experiment

Decoding Accuracy

			5th	25th	50th	Mean
Webcam	Printer	Enterprise	88%	94%	98%	95.9%
		Consumer	90%	98%	99%	98.1%
		Pro	97%	99%	100%	99.2%
	Screen	Monitor	94%	98%	99%	98.5%
		Laptop	97%	99%	100%	99.1%
		Cellphone	91%	98%	99%	97.7%
Cellphone	Printer	Enterprise	88%	96%	98%	96.8%
		Consumer	95%	99%	100%	99.0%
		Pro	97%	99%	100%	99.3%
	Screen	Monitor	98%	99%	100%	99.4%
		Laptop	98%	99%	100%	99.7%
		Cellphone	96%	99%	100%	99.2%
DSLR	Printer	Enterprise	86%	96%	99%	97.0%
		Consumer	97%	99%	100%	99.3%
		Pro	98%	99%	100%	99.5%
	Screen	Monitor	99%	100%	100%	99.8%
		Laptop	99%	100%	100%	99.8%
		Cellphone	99%	100%	100%	99.8%

Encoded images for different message length

- To embed longer message, some square artifacts appear

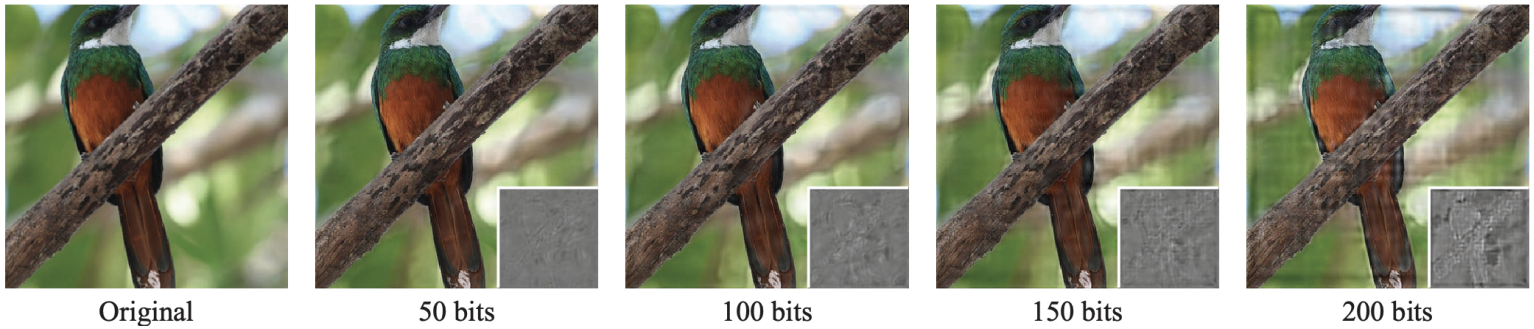


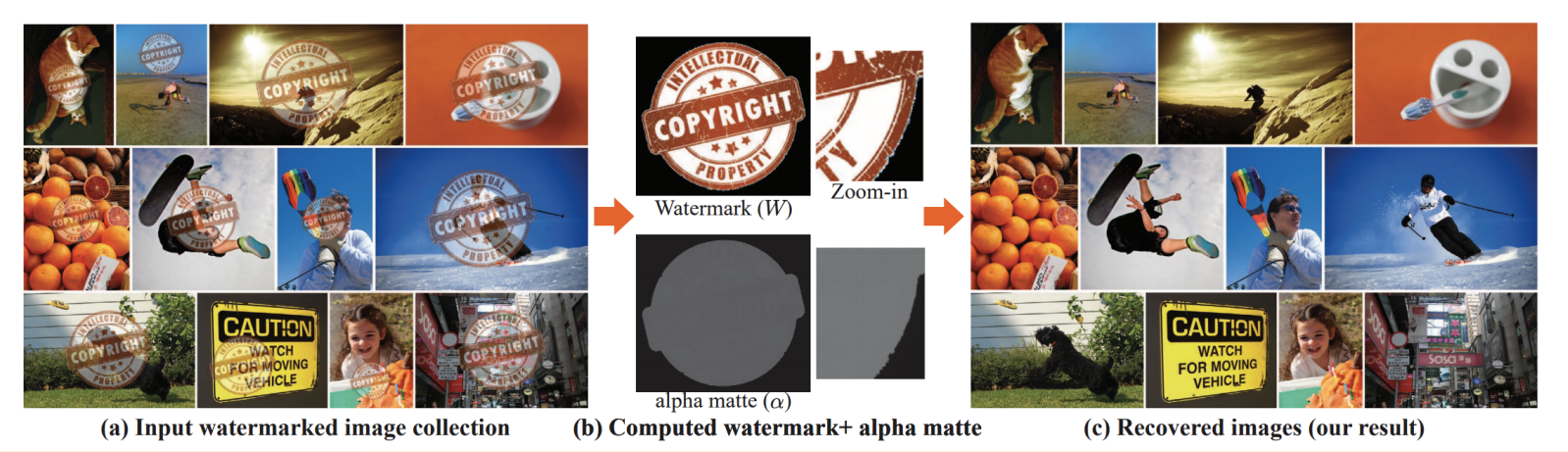
Figure 8: Four models trained to encode messages of different lengths. The inset shows the residual relative to the original image. The perceptual quality decreases as more bits are encoded. We find that a message length of 100 bits provides good image quality and is sufficient to encode a virtually unlimited number of distinct hyperlinks using error correcting codes.

Visible Watermark Removal

Is it really effective?

On the Effectiveness of Visible Watermarks

- Easy to remove visible watermark by exploiting collection of watermarked images
- This is image collection approach, it's new

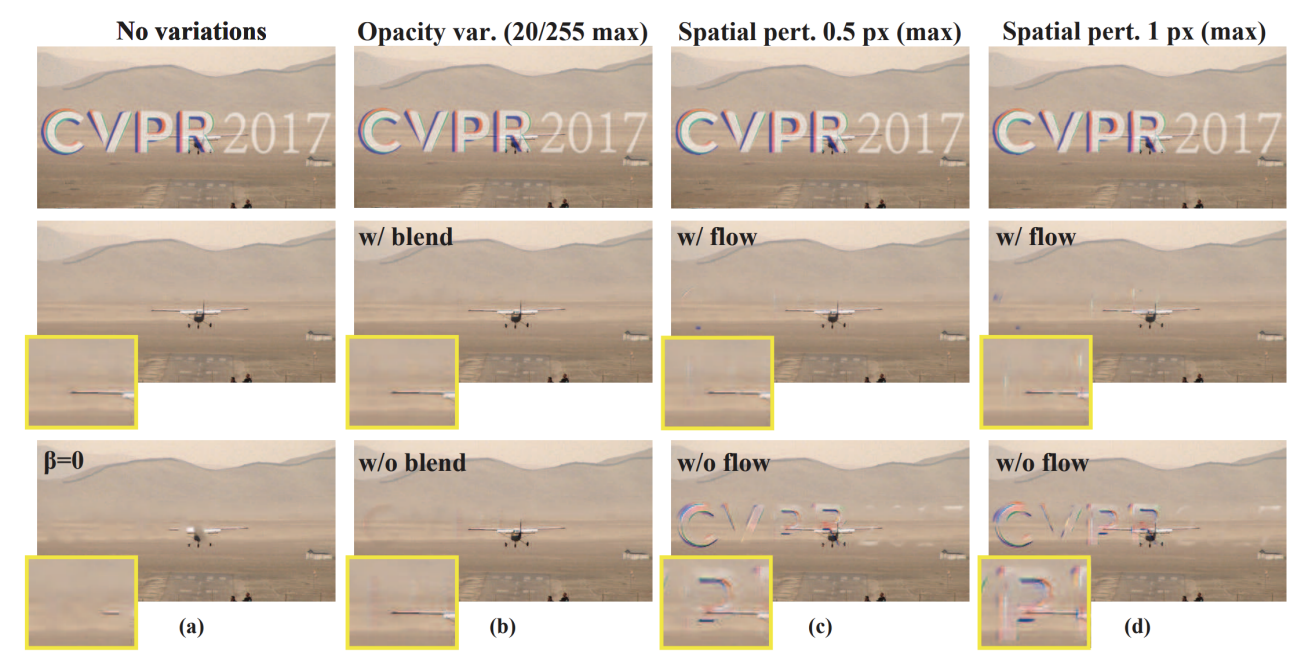


Comparison to baseline

- Our result is best quality

Analysis for inconsistency watermark pattern

- Not robust to spatial perturbation
- Do perturbation affect to reconstruction phase?



Result for stock imagery

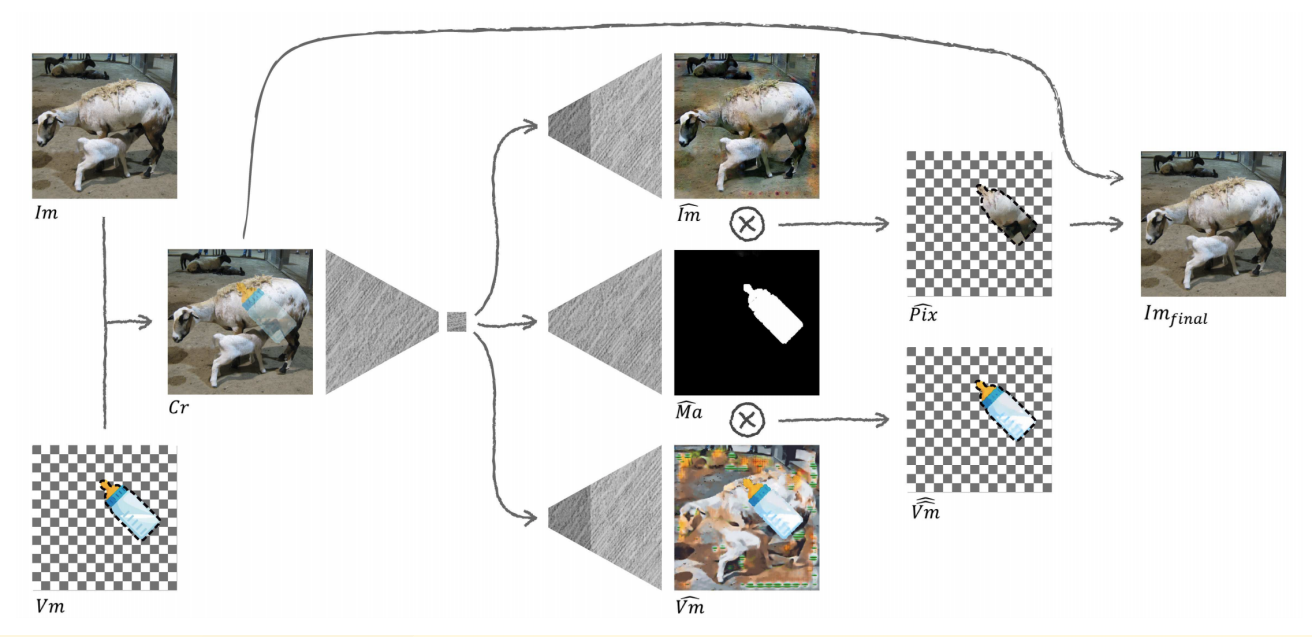
Datasets	# Images	Long Edge	Watermak Res.(w×h)	Pipeline (min)	Removal (sec)
AdobeStock	422	1000	623 × 134	33	18
123RF	1376	650	650 × 433	115	60
CanStock	3000	450	450 × 300	28	12
fotolia	285	500	199 × 66	5	1.5
CVPR17	1000	640	403 × 67	34	5
Copyrights	1000	640	339 × 307	47	40

Blind Visual Motif Removal from a Single Image

- Remove visual motif (such as watermark) from image by using only single image
- Easy to remove
- This is blind removal

Method

- Extract background, motif area, foreground (motif) from the image, and then reconstruct the original image



Result images

Appendix: Model Watermarking

Protect copyrights of you model!

Coming soon!

Summary

Invisible Watermark with NN

- Under development
- Encoder / Adversary / Decoder framework is the best
- Robustness is most important

Visible Watermark Removal

- Single Image VS Multiple Image (Blind VS Non-Blind)
- Strong watermark which cannot be removed is needed