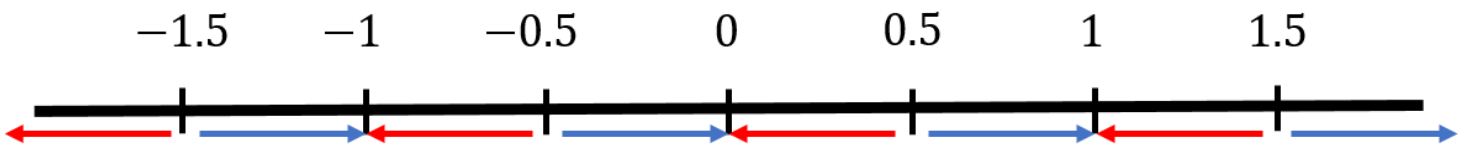
Steganography and its Detection in JPEG Images Obtained with the "Trunc" Quantizer

Jan Butora and Jessica Fridrich

ICASSP 2020

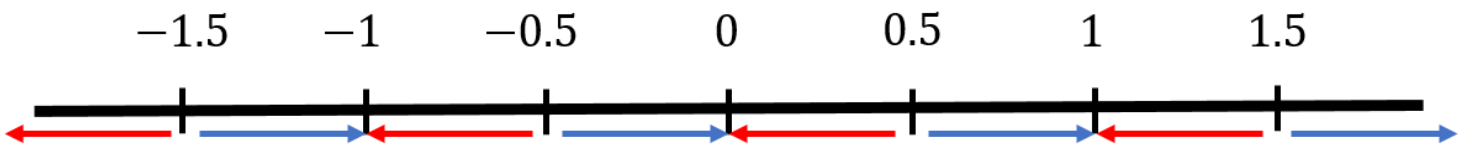
BINGHAMTON
UNIVERSITY
STATE UNIVERSITY OF NEW YORK

What is Trunc Quantizer?

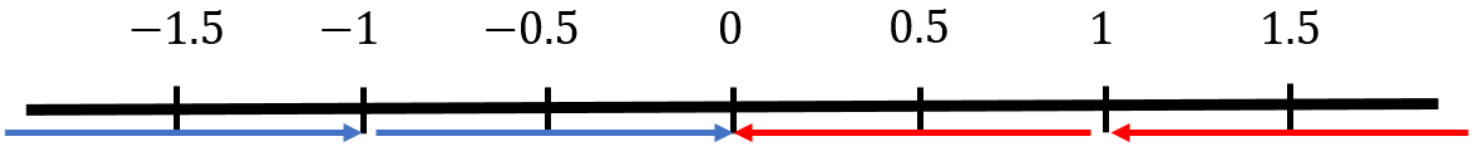


Round quantizer

What is Trunc Quantizer?



Round quantizer

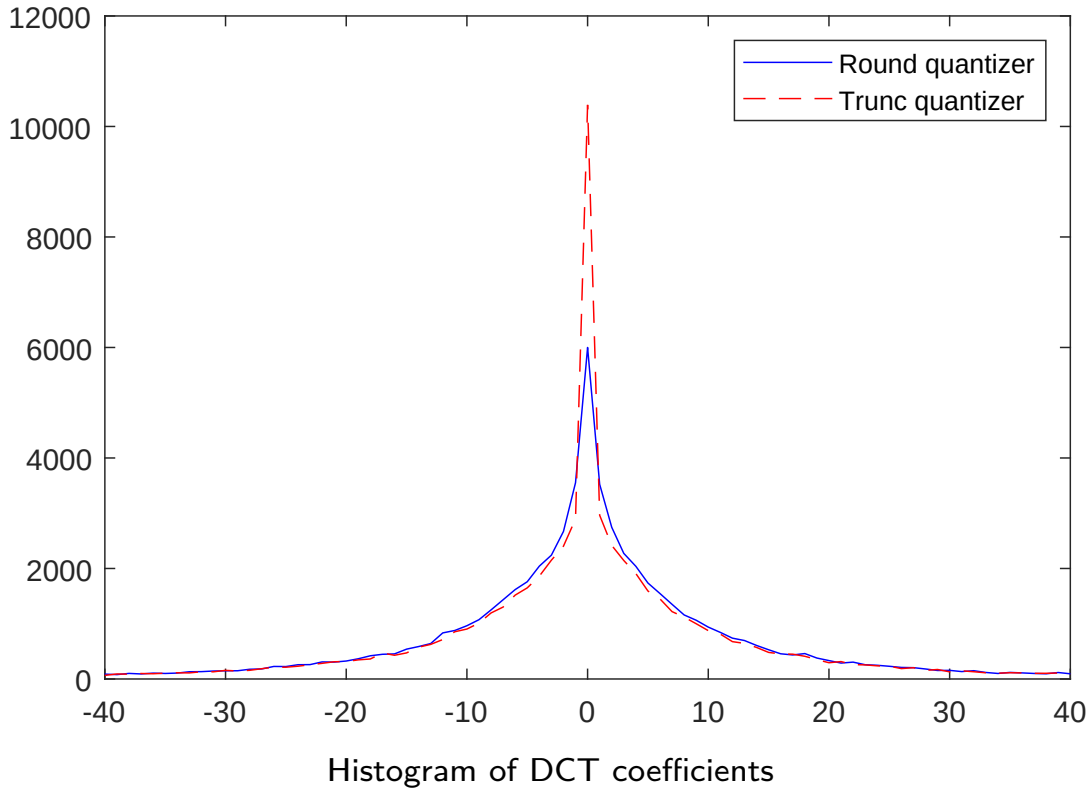


Trunc quantizer

- easier to implement in hardware
- Very common: iPhone 5c, Canon EOS 10D, Samsung Galaxy Tab 3 8.0

Effect of Quantizer on Histogram of DCTs

- One image compressed with quality factor 100 with round and trunc quantizers



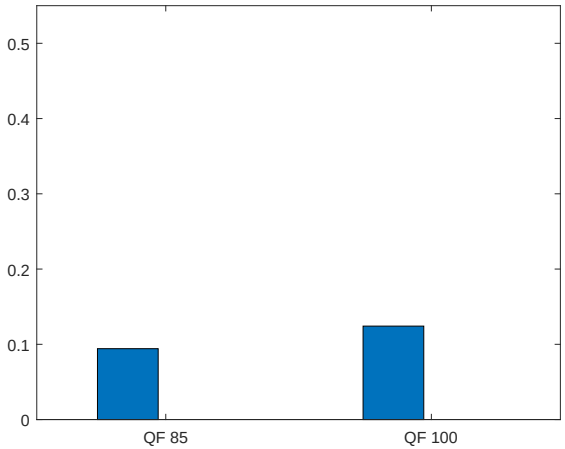
Trunc JPEGs and Steganography

Existence of trunc JPEGs has serious implications both the Steganographer and the Steganalyst:

- Some steganographic schemes (J-UNIWARD, SI-UNIWARD) need to be redesigned to prevent security holes
- Steganalysis also needs to be redesigned for best results

Unaware Steganalyst completely fails

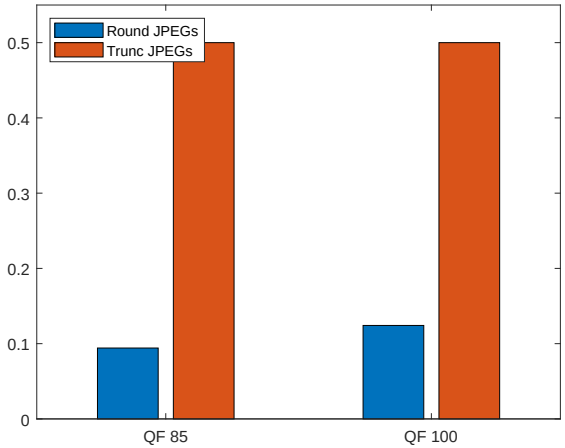
- Training on round JPEGs and testing on trunc JPEGs leads to catastrophic detection failure



Total detection error under equal priors P_{E} of SRNet on J-UNIWARD, 0.4 bpnzac

Unaware Steganalyst completely fails

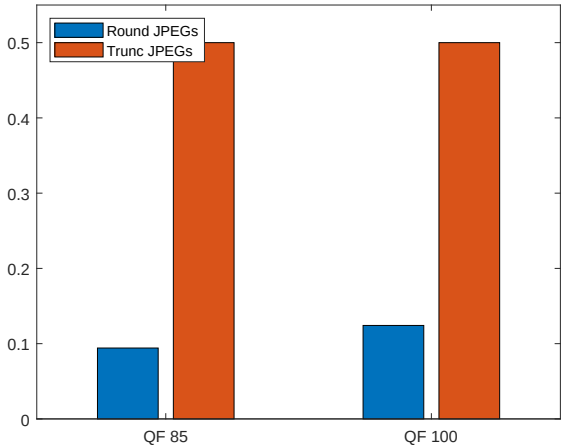
- Training on round JPEGs and testing on trunc JPEGs leads to catastrophic detection failure



Total detection error under equal priors P_{E} of SRNet on J-UNIWARD, 0.4 bpnzac

Unaware Steganalyst completely fails

- Training on round JPEGs and testing on trunc JPEGs leads to catastrophic detection failure



Total detection error under equal priors P_{E} of SRNet on J-UNIWARD, 0.4 bpnzac

Failure common across QFs, stego algorithms, detectors

- QFs 75, 76, ..., 100
- SRNet, JRM, GFR
- J-UNIWARD, nsF5, UED
- P_{FA} 99–100%

Experimental Setup

- Dataset: BOSSbase 1.01 + BOWS2, 20,000 grayscale images resized to 256×256
  - Round JPEGs
  - Trunc JPEGs
- TRN / VAL / TST: 10,000BOWS2 + 4,000BOSS / 1,000BOSS / 5,000BOSS
- Stego algorithms: nsF5 (0.2 bpnzac), UED (0.3 bpnzac), J-UNIWARD (0.4 bpnzac)
  - all assumed optimally coded (embedding simulator)

Evaluation Metric

- Accuracy $= 1 - P_{\mathrm{E}}$
  - Total classification error $P_{\mathrm{E}} = \frac{1}{2}(P_{\mathrm{FA}} + P_{\mathrm{MD}})$
  - P_{FA} false alarm rate
  - P_{MD} missed detection
- Detectors
  - SRNet [Boroumand 2018]
  - GFR (Gabor Filter Residual) feature set coupled with ensemble classifier
  - JRM (JPEG Rich Model) feature set coupled with ensemble classifier

Trunc and Round JPEGs can be reliably distinguished

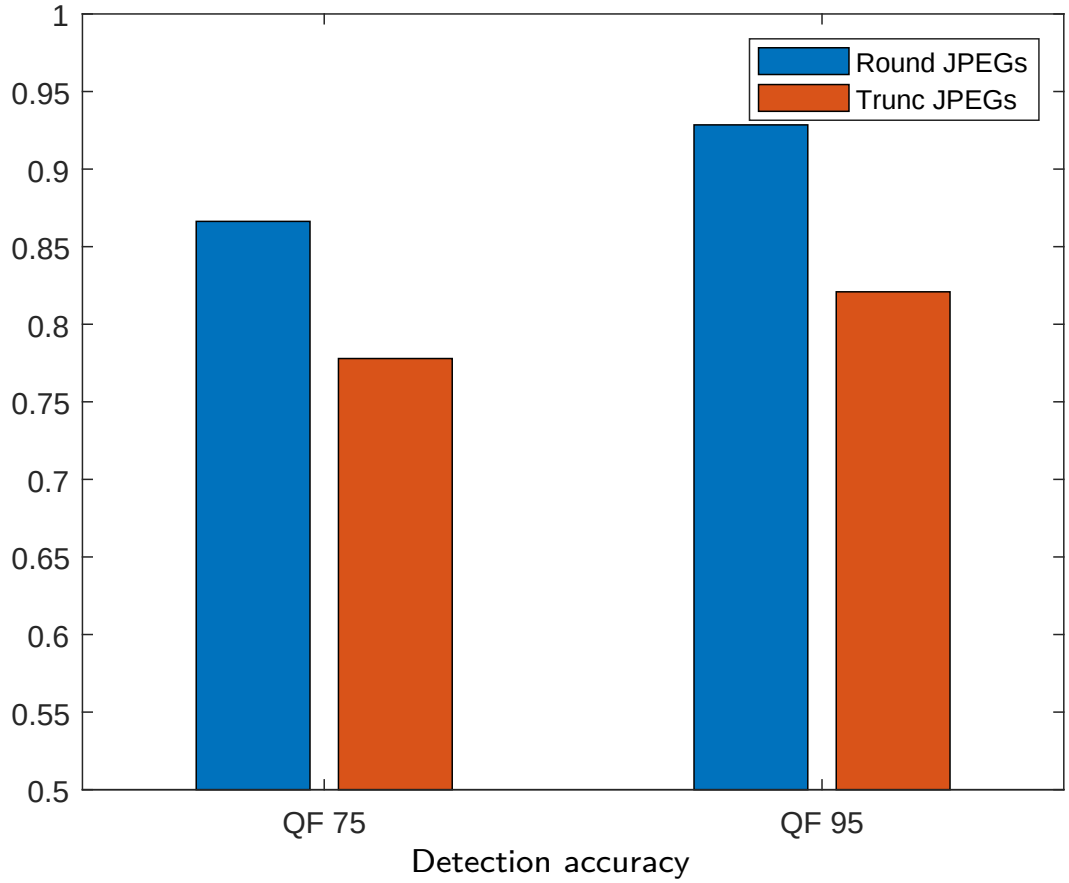
- Train SRNet between Round **covers** and Trunc **covers**
- Accuracy very close to 100%
  - For quality factors 85, 100
  - Even when tested on stego images
- $\Longrightarrow$ one can build separate detectors for each cover source

Effect of Trunc Quantizer on Security

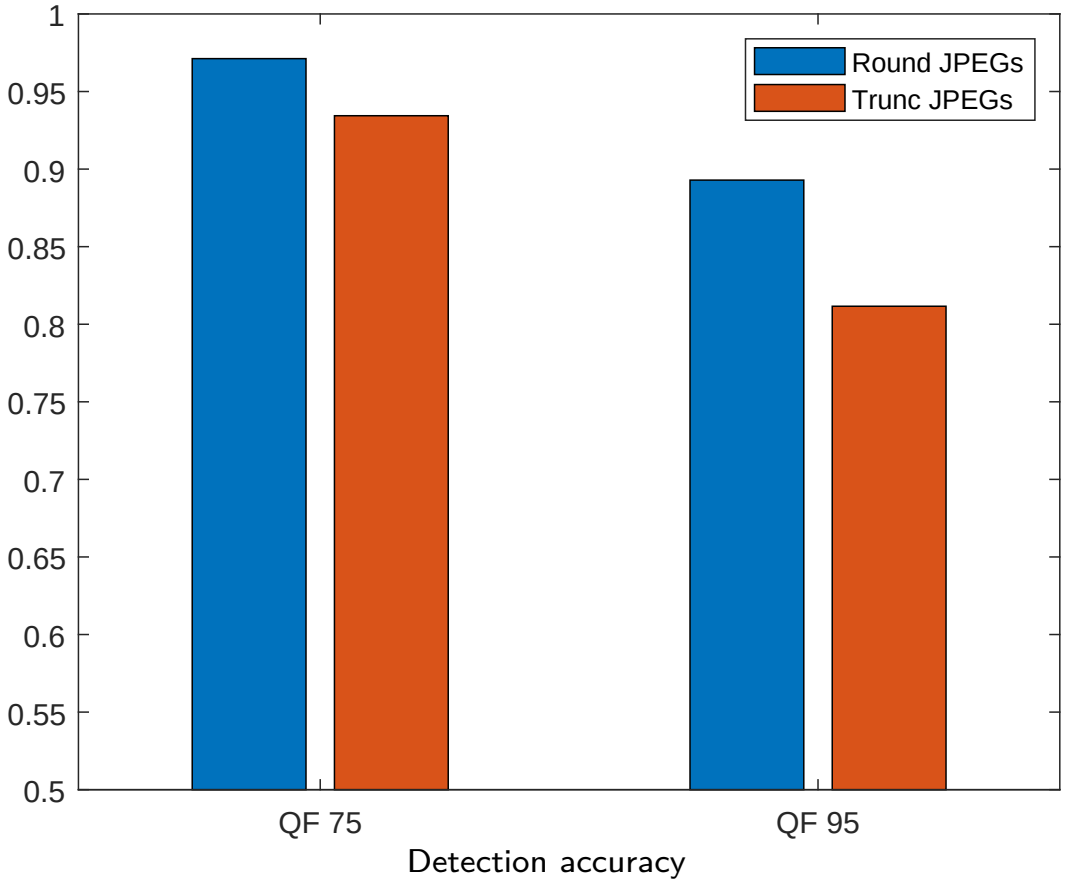
Is it harder or easier to detect stego in round or trunc JPEGs?

- Trunc JPEGs have more zero DCT coefficients
- Thus, for fair comparison, payload for trunc JPEGs was scaled according to Square Root Law
- Conclusions similar for fixed bpnzac and bpp

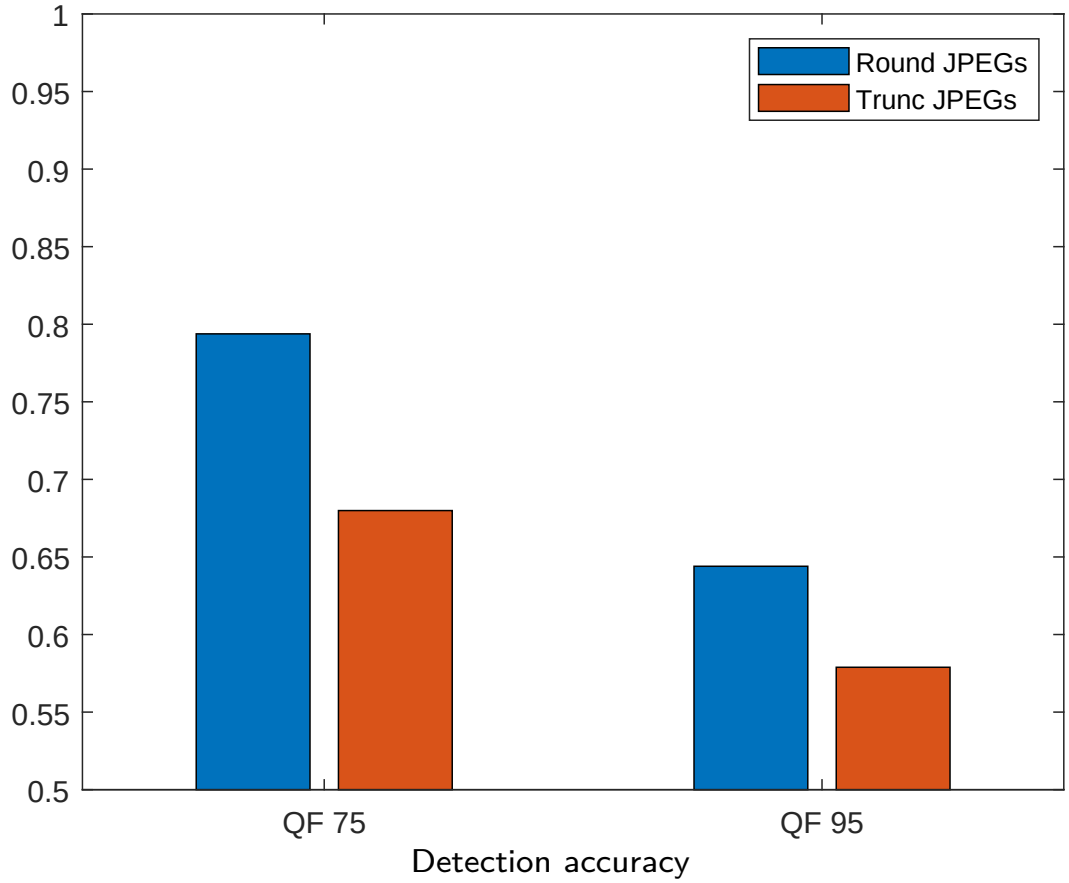
nsF5, 0.2 bpnzac - JRM



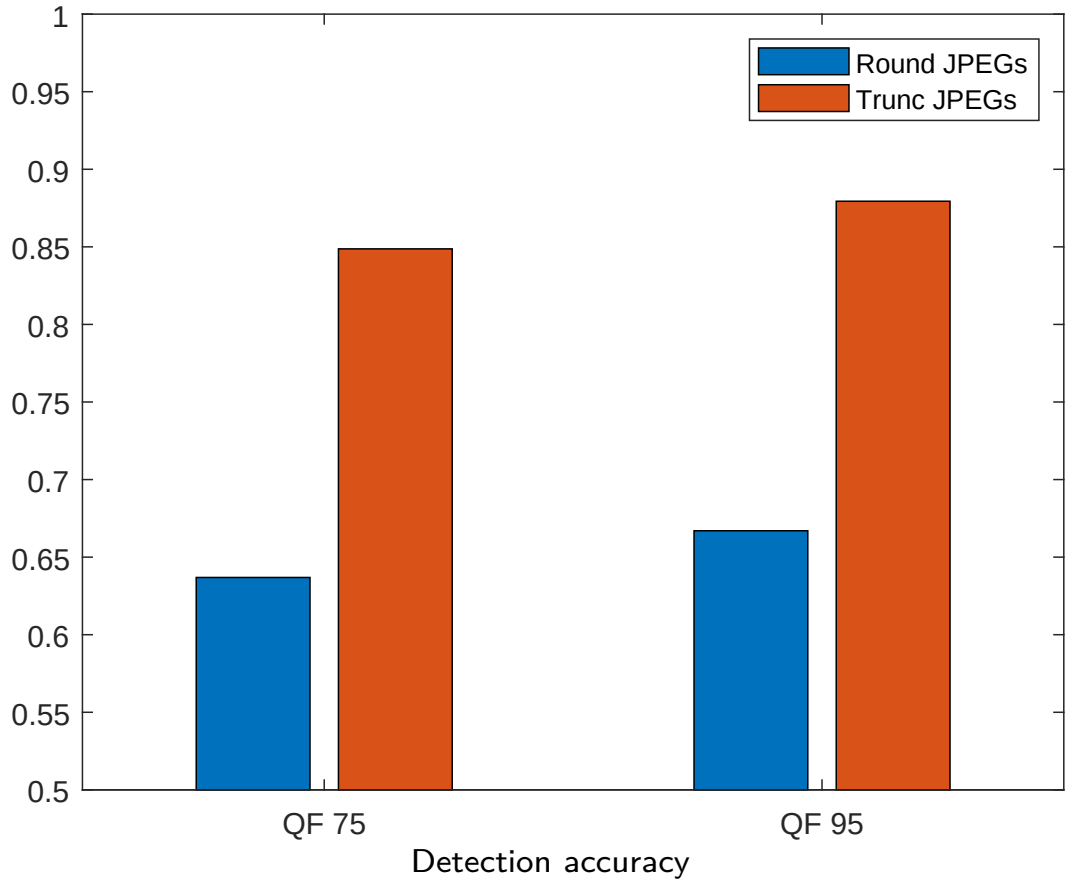
UED, 0.3 bpnzac - SRNet



J-UNIWARD, 0.4 bpnzac - GFR



J-UNIWARD, 0.4 bpnzac - JRM



Wait, what??

Why is J-UNIWARD so detectable in Trunc JPEGs with JRM?

- J-UNIWARD is the only tested algorithm that embeds into zero coefficients
- There are many more zero coefficients in trunc JPEGs

Why is J-UNIWARD so detectable in Trunc JPEGs with JRM?

- J-UNIWARD is the only tested algorithm that embeds into zero coefficients
- There are many more zero coefficients in trunc JPEGs

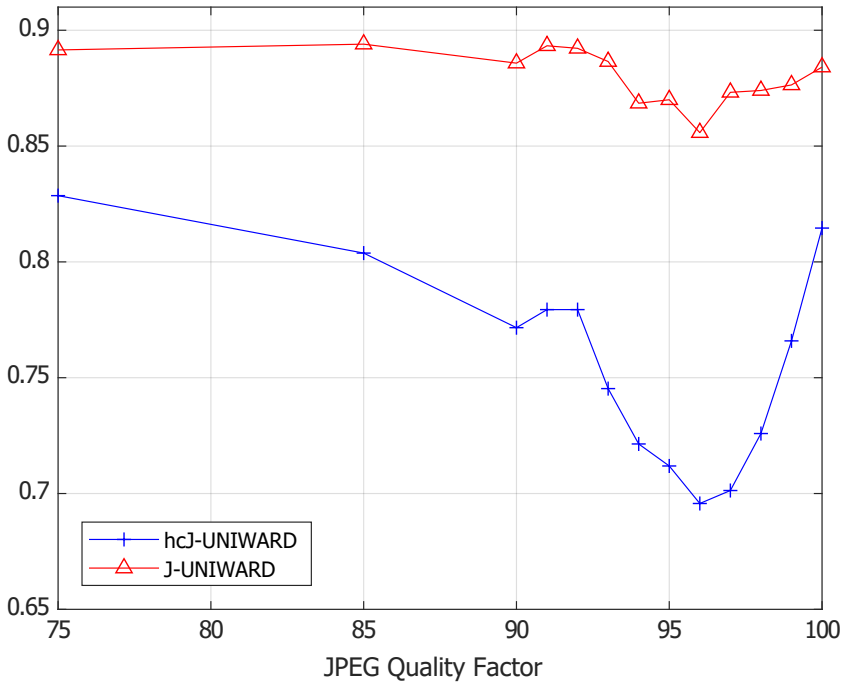
J-UNIWARD embeds **too much** into zero coefficients in trunc JPEGs!

J-UNIWARD for trunc JPEGs

- Change embedding costs of zero coefficients, so that on average the number of zero coefficients is preserved
  - hcJ-UNIWARD (J-UNIWARD with **h**istogram **c**orrection)
- Achieved by increasing the cost of zeros $\rho_0 \to \tilde{\rho}_0 = \eta\rho_0$ by factor $\eta = \frac{\rho_1}{\rho_0} + \frac{1}{\lambda\rho_0} \log\left(\frac{2h[0]}{h[1]+h[-1]}\right)$ (details in the paper)

hcJ-UNIWARD vs J-UNIWARD

- Three detectors: SRNet, JRM, union of SRNet features (512-dim input to IP layer) concatenated with JRM features coupled with ensemble classifier



Accuracy of the best detector in trunc JPEGs at 0.4 bpnzac

Side information

- Heuristic side-informed schemes for round JPEGs cannot be used
- Need to take into account
  - different range of rounding errors, $0 \leq e < 1$ for positive DCTs
  - increased number of zero coefficients

Side information

- Heuristic side-informed schemes for round JPEGs cannot be used
- Need to take into account
  - different range of rounding errors, $0 \leq e < 1$ for positive DCTs
  - increased number of zero coefficients
- Proposed SI-UNIWARD for trunc JPEGs
  - Minimum-perturbation modulation (see paper)
  - No need for histogram correction

bpnzac	1	0.8	0.6	0.4
QF75	0.8164	0.7436	0.6485	0.5653
QF95	0.7984	0.6972	0.6050	0.5420

Detection accuracy with SRNet

Conclusions

- Steganalyst unaware of trunc JPEGs will expereince 100% false alarm
- Easy fix by training a detector for each source (detecting quantizer type is reliable)
- Trunc JPEGs are more friendly for steganographers
  - algorithms that embed into zero coefficients need to be adjusted by increasing the cost of modifying zeros
- Redesigned side-informed schemes to take into account different range of rounding errors