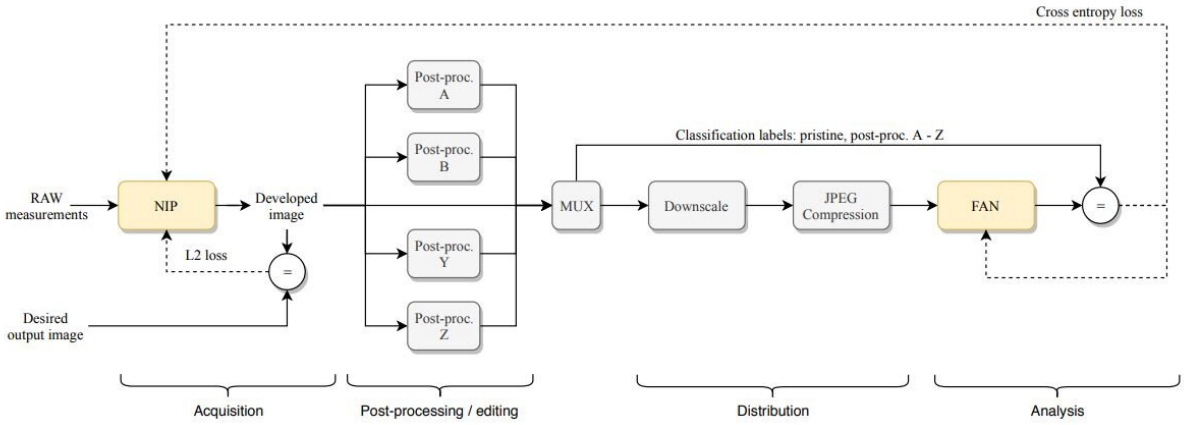
Nailing digital fakes with AI-learned artifacts

May 30 2019, by Nancy Cohen



Optimization of the image acquisition and distribution channel to facilitate photo provenance analysis. The neural imaging pipeline (NIP) is trained to develop images that both resemble the desired target images, but also retain meaningful forensic clues at the end of complex distribution channels. Credit: arXiv:1902.10707 [cs.CV] https://arxiv.org/abs/1902.10707

We see the imaginative feats of photo fakery; now we have to figure out what to do about them. Being able to tell fake from real is the goal, but how to get there? Forensics is the key tool to hunt down fake photos and it does not appear to be an easy task in getting that tool to perform well.

"One of the most difficult things about detecting manipulated photos, or 'deepfakes,' is that digital photo files aren't coded to be tamper-evident,"

said Lily Hay Newman in *Wired*.

What have the experts achieved, then? Forensic analysts figured out how to spot some digital characteristics to detect meddling, "but these indicators don't always paint a reliable picture," she said.

And even those clues may not help, as "many common types of 'post-processing,' like file compression for uploading and sharing photos online, strip away these clues anyway."

But hold on. A New York University Tandon School of Engineering pair of researchers had a what-if —a tamper-resistant seal from the camera itself.

Their paper discussing this idea is up on arXiv, and it is titled, "Neural Imaging Pipelines—the Scourge or Hope of Forensics?" The authors are Pawel Korus and Nasir Memon.

"We demonstrate that a neural network can be trained to replace the entire photo development pipeline, and jointly optimized for high-fidelity photo rendering and reliable provenance analysis. Such optimized neural imaging pipeline allowed us to increase image manipulation detection accuracy from approx. 45% to over 90%. The network learns to introduce carefully crafted artifacts, akin to digital watermarks, which facilitate subsequent manipulation detection. Analysis of performance trade-offs indicates that most of the gains can be obtained with only minor distortion."

Wired explained what the authors proposed: training a neural network to power the photo development process that takes place inside the cameras. "The sensors are interpreting the light hitting the lens and turning it into a high quality image, the neural network is also trained to mark the file with indelible indicators that can be checked later, if

needed, by forensic analysts," Newman wrote.

She quoted researcher Nasir Memon commenting on checking for fakes in this manner. He said that "you have to go close to the source where the image is captured."

He further said that in this work "we are creating an image which is forensics-friendly, which will allow better forensic analysis than a typical image. It's a proactive approach rather than just creating images for their visual quality and then hoping that forensics techniques work after the fact."

Melanie Ehrenkranz in *Gizmodo* also clarified what the researchers were trying to accomplish, to gain success in forensics using machine learning for forensics purposes, and baking a detection method right into the camera.

Ehrencranz: "They detail a method in which a neural network replaces the photo development process so that the original image taken is marked with something like a digital watermark to indicate the photo's provenance in a digital forensics analysis. In other words, the process identifies a photos origin and whether it has been manipulated since its original state."

The news release from the NYU Tandon School of Engineering had an especially good summary of what these researchers have achieved. Their approach "replaces the typical photo development pipeline with a neural network—one form of AI—that introduces carefully crafted artifacts directly into the image at the moment of image acquisition. These artifacts, akin to 'digital watermarks,' are extremely sensitive to manipulation."

"Unlike previously used watermarking techniques, these AI-learned

artifacts can reveal not only the existence of photo manipulations, but also their character," Korus said.

The process is optimized for in-camera embedding and can survive image distortion applied by online photo sharing services.

The discussion has been about forensic watermarking on photos. What about video? *Wired* said video was something the researchers said they did not broach yet, but that it would be theoretically possible.

"We believe it is imperative to consider new opportunities for security-oriented design of cameras and multimedia dissemination channels that come with adoption of neural imaging processors."

Actually, their neural imaging toolbox is available at GitHub. It's described as a "Python toolbox for optimization of neural imaging pipelines for photo manipulation detection."

The NYU Tandon release made the point that in the coming years, "AI-driven processes are likely to fully replace the traditional digital imaging pipelines." Memon said that as this transition takes place, 'we have the opportunity to dramatically change the capabilities of next-generation devices when it comes to image integrity and authentication. Imaging pipelines that are optimized for forensics could help restore an element of trust in areas where the line between real and fake can be difficult to draw with confidence.'"

More information: Neural Imaging Pipelines - the Scourge or Hope of Forensics? arXiv:1902.10707 [cs.CV] arxiv.org/abs/1902.10707