Linking Research to the Real World

A look at what's driving research and where it's going

Anurag Malyala

2K15/CO/035

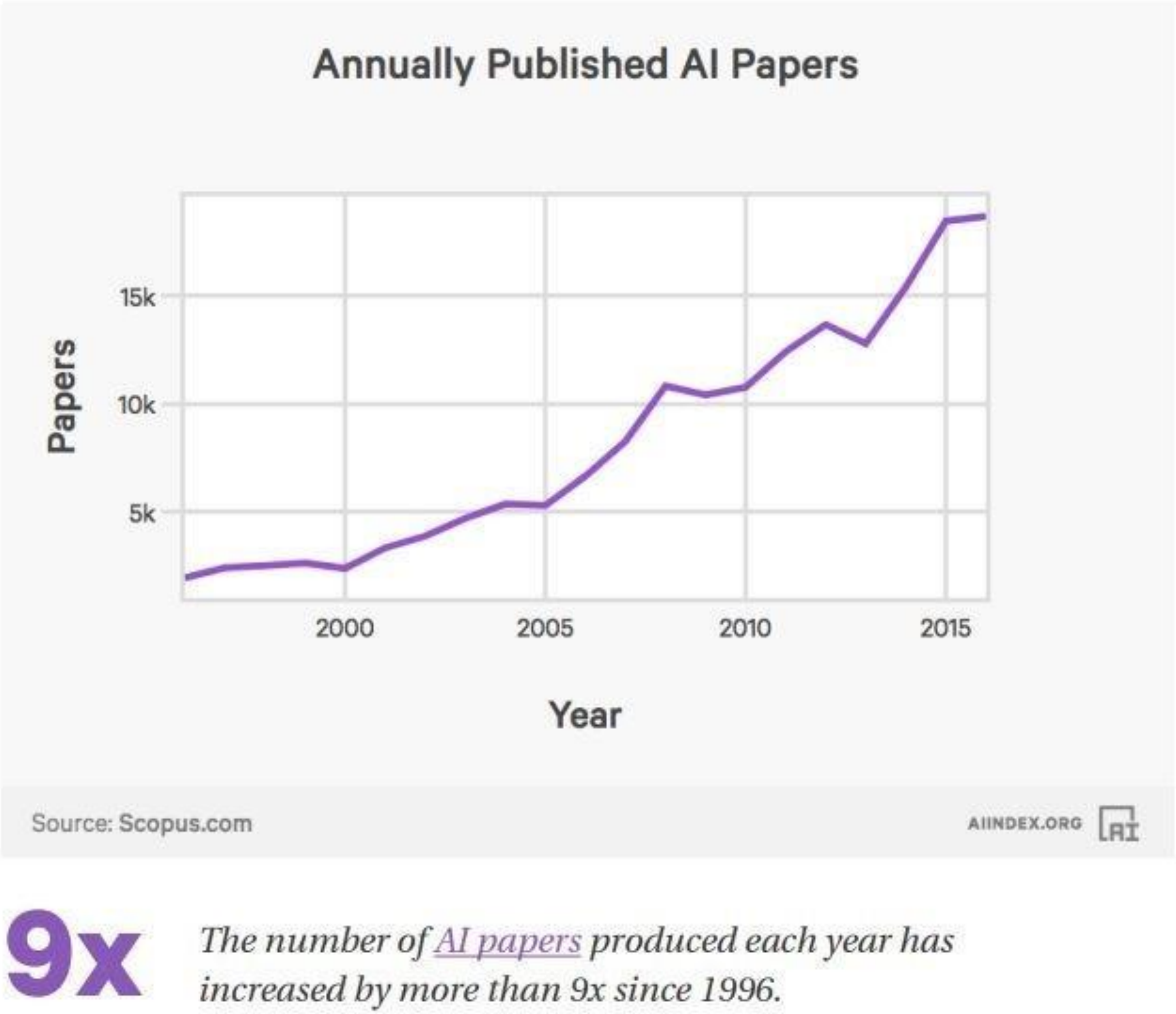
CO-403, Seminar

Introduction

State of AI Research today

State of AI Research

- Publications featuring AI have seen a sharp increase in recent years.
- This can be attributed to the cheap and easy availability of High-Performance Computing Systems through Democratization of the Cloud Computing and GPU platforms.
- This has widened the disconnect between state of art in research and the products and services.



State of AI Research

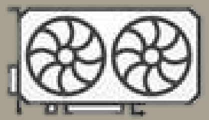
- The disconnect is not always there though.
- Computer Aided Medicine and Image processing have shown to keep pace with the research in respective field.
- AI enabled diagnosis and drug discovery systems are at par if not beyond human performance.
- AI methods combined with smart signal processing methods have revolutionized photography and image processing tasks, bringing tools once only available to professional or ones that were only known by a few to the masses

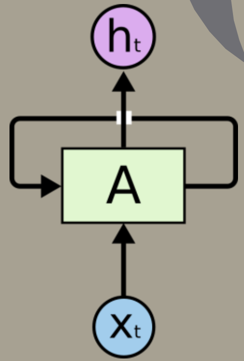
State of AI Research

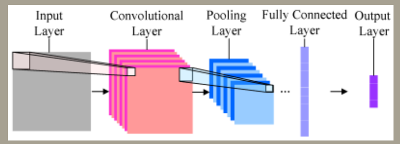
- Other domains which have begun to shorten this disconnect are Natural Language systems.
- Voice assistants have pushed the product sector to catch up with research.
- Google’s Duplex technology is the state of art the voice generation and has started to roll out to consumer devices as Google Call Screen.

The Fire Triangle

The three things that have accelerated research and helped closing the disconnect

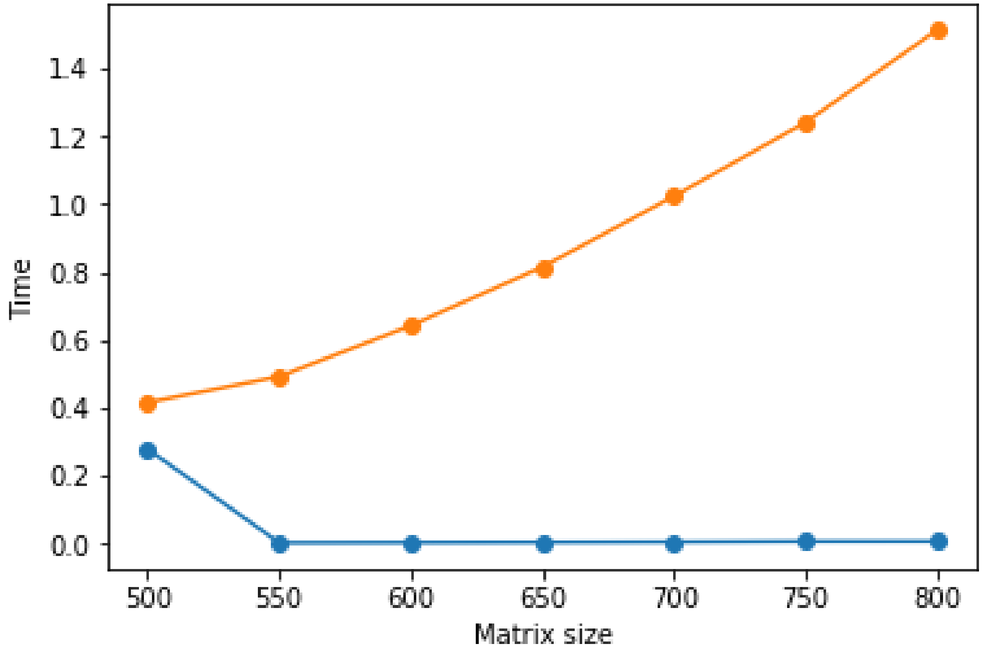






Graphic Processing Units

- Nearly all data is represented as numbers in form of Vectors and Matrices.
- Operations on matrices are computationally expensive but highly parallelizable
- GPUs enable for high bandwidth highly parallel computing capabilities offering processing cores 10-100x more than the best CPUs



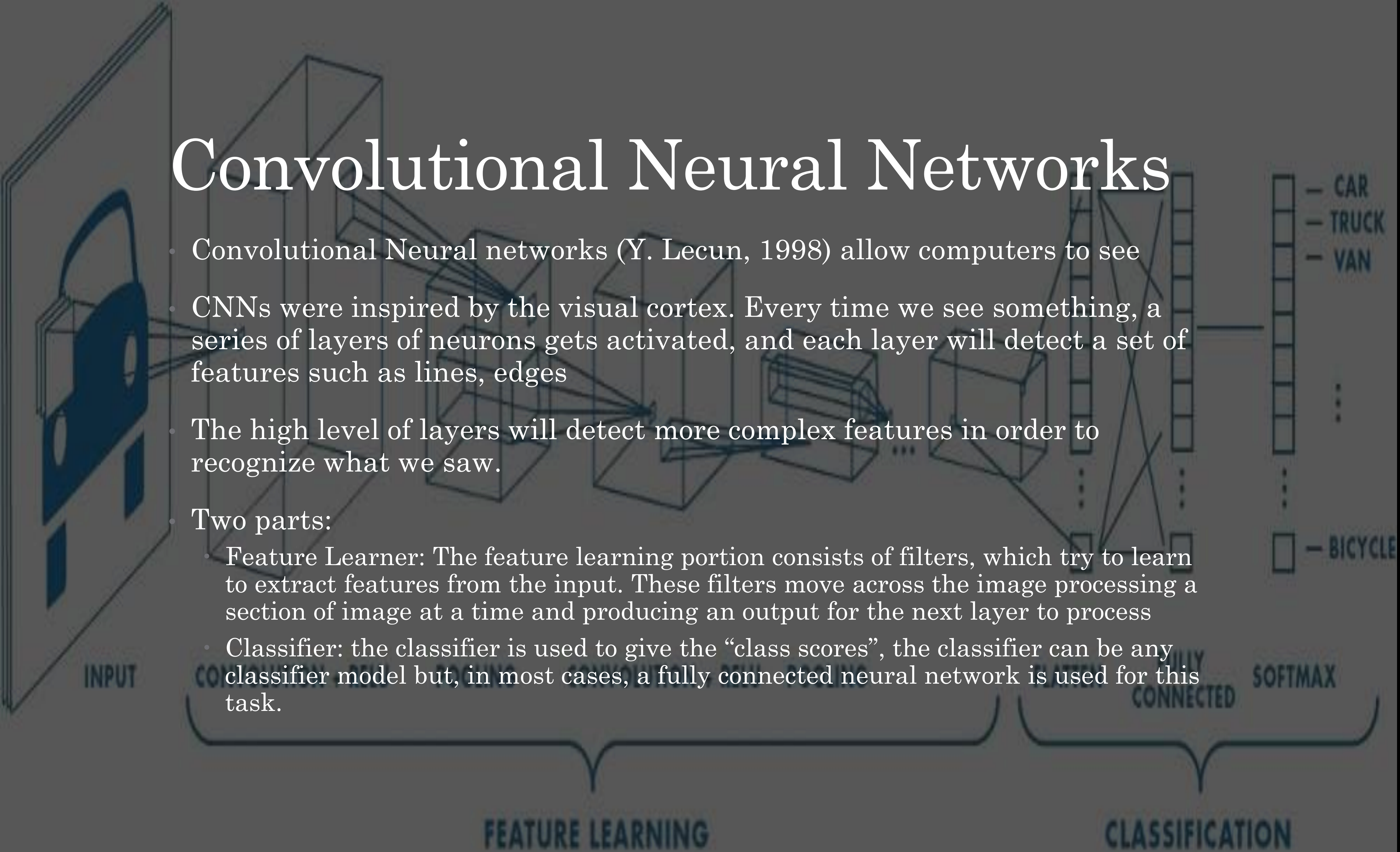
CPU vs GPU, Increase in computation time with increase in input size of matrix. Comparing Nvidia GTX 1080 with latest Generation Inter Core i7

Graphic Processing Units

- Packages in terms of software and hardware from manufactures have made accessing and using GPU easier.
- PyTorch, TensorFlow, are two of the most common software packages that provide a high-level API for working on GPU
- Nvidia with their NVCloud and partnerships with Google and Amazon has made swaths of GPUs available to public on Cloud Platforms along with their CUDA Runtime.

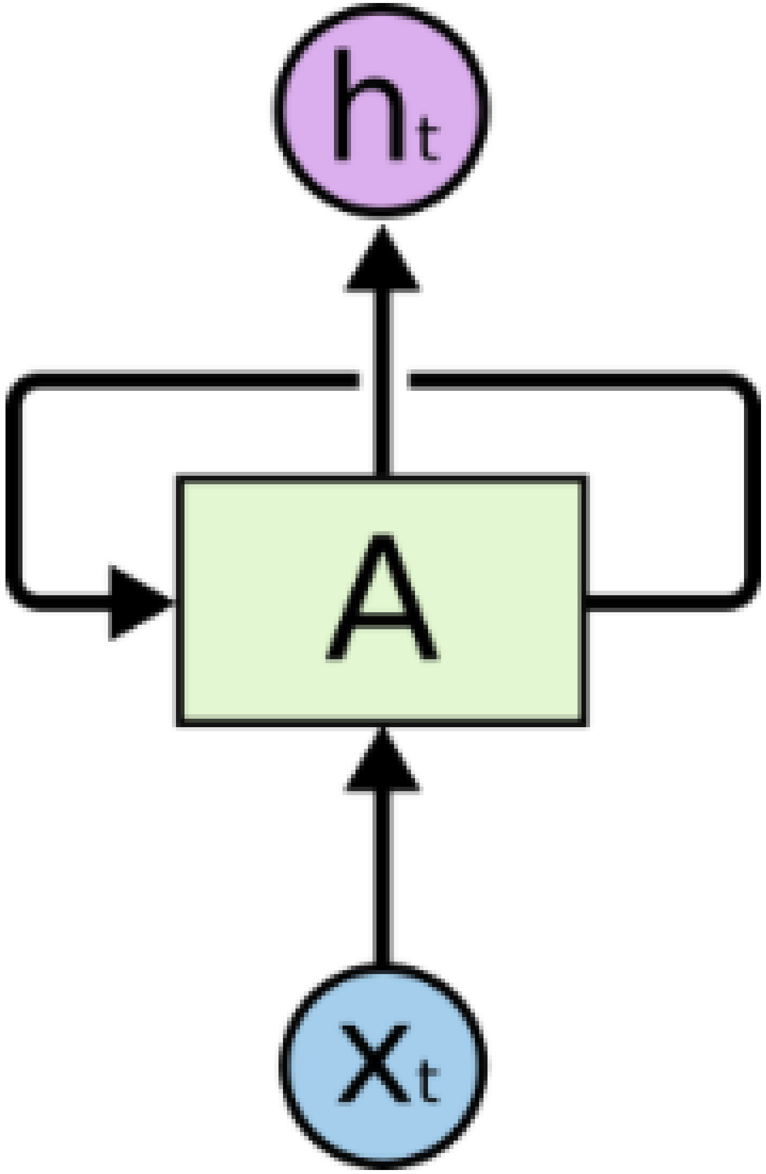
Convolutional Neural Networks

- Convolutional Neural networks (Y. Lecun, 1998) allow computers to see
- CNNs were inspired by the visual cortex. Every time we see something, a series of layers of neurons gets activated, and each layer will detect a set of features such as lines, edges
- The high level of layers will detect more complex features in order to recognize what we saw.
- Two parts:
  - Feature Learner: The feature learning portion consists of filters, which try to learn to extract features from the input. These filters move across the image processing a section of image at a time and producing an output for the next layer to process
  - Classifier: the classifier is used to give the “class scores”, the classifier can be any classifier model but, in most cases, a fully connected neural network is used for this task.



Convolutional Neural Networks

- The entire CNN can be summarized as follows:
  - Provide input image into convolution layer
  - Choose parameters, apply filters with strides, padding if requires. Perform convolution on the image and apply non-linearity to the matrix.
  - Perform pooling to reduce dimensionality size
  - Add as many convolutional layers until satisfied
  - Flatten the output and feed into a fully connected layer (FC Layer)
  - Output the class using an activation function and classifies images



Recurrent Neural Networks

- While CNNs are the state of art for images, for speech*, time stamped data, Recurrent networks are king.
- Unlike feedforward neural networks, RNNs can use their internal state (memory) to process sequences of inputs. This makes them applicable to tasks such as unsegmented, connected handwriting recognition or speech recognition.
- The current output along with the next time steps input are combined to make the next input for the network

Recurrent Neural Networks

- The recurrent nature of processing allows them to develop a “memory” element.
- When dealing with word sequences, the semantics of the sentence are preserved and not forgotten, when the next word is processed, there is seamless flow the entire semantic history for processing.
- This allows for the network to develop long-term-dependencies.
- *With WaveNet DeepMind stole the state of art crown in speech generation from RNNs in favor of CNNs.

Two domains at the bleeding edge

A showcase of cases where the real world is at level with research

Medicine

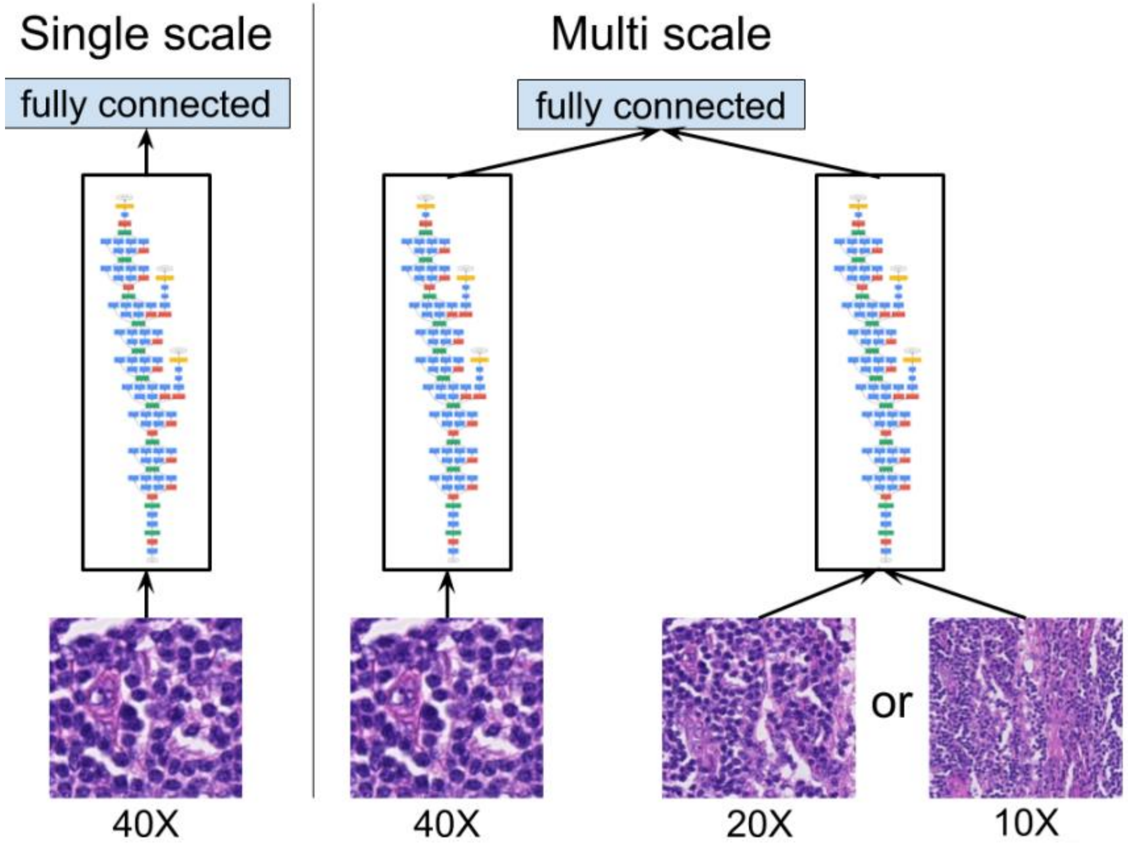
Diagnosis and Drug Discovery

Applying Deep Learning to Metastatic Breast Cancer Detection

- In October 2018, Google published the results of their work in diagnosing Metastic Breast Cancer, claiming an accuracy of 99%.
- Such a high accuracy score is in fact better than the current gold standard, a pathologist with a microscope.
- This can be traced back to a 2017 paper, “Detecting Cancer Metastases on Gigapixel Pathology Images”
- This used a convolutional neural network (CNN) approach for segmenting gigapixel pathology images into normal and cancerous pixels to aid breast cancer diagnosis
- Giga-pixel images are of the order $10^6 x 10^6$ making all other methods unviable.

Applying Deep Learning to Metastatic Breast Cancer Detection

- They used a modified CNN architecture called Inception.
- Using both multiscale images and data augmentation

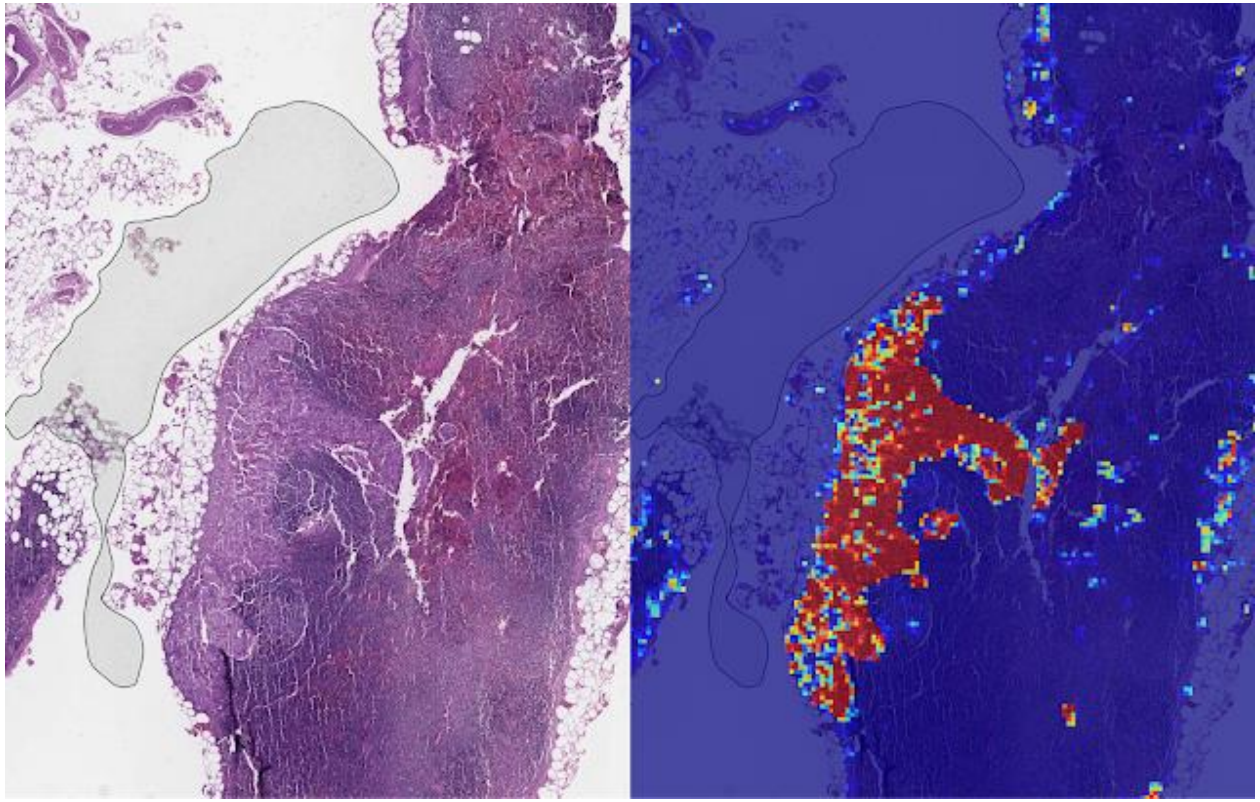


Applying Deep Learning to Metastatic Breast Cancer Detection

- The automated approach process the scans in sections of size $300x300$ and can detect cancer cells as small as $100x100$
- They proposed:
  - For each input image, they chose to process the image in patches of 299 by 299 for each image the classified them as containing a cancerous tumor or a normal cell around a central region of 128 by 128
  - They label a patch as cancerous if at least one of the pixels in the center regions is labeled cancerous
- This preliminary work gave an accuracy of 92.4% with a large margin for improvement where as a human pathologist could only achieve 73.2%.

Applying Deep Learning to Metastatic Breast Cancer Detection

- Results



Left: sample view of a slide containing lymph nodes, with multiple artifacts: the dark zone on the left is an air bubble, the white streaks are cutting artifacts, the red hue across some regions are haemorrhagic (containing blood), the tissue is necrotic (decaying), and the processing quality was poor.

Right: LYNA identifies the tumor region in the center (red), and correctly classifies the surrounding artifact-laden regions as non-tumor (blue).

Automating Drug Discovery

- Protein crystallization is a key step to biomedical research concerned with discovering the structure of complex biomolecules.
- Because that structure determines the molecule's function, it helps scientists design new drugs that are specifically targeted to that function.
- Protein crystals are rare and difficult to find.
- Setup and Imaging processes have been automated but the task of recognizing the proteins is largely done by humans.

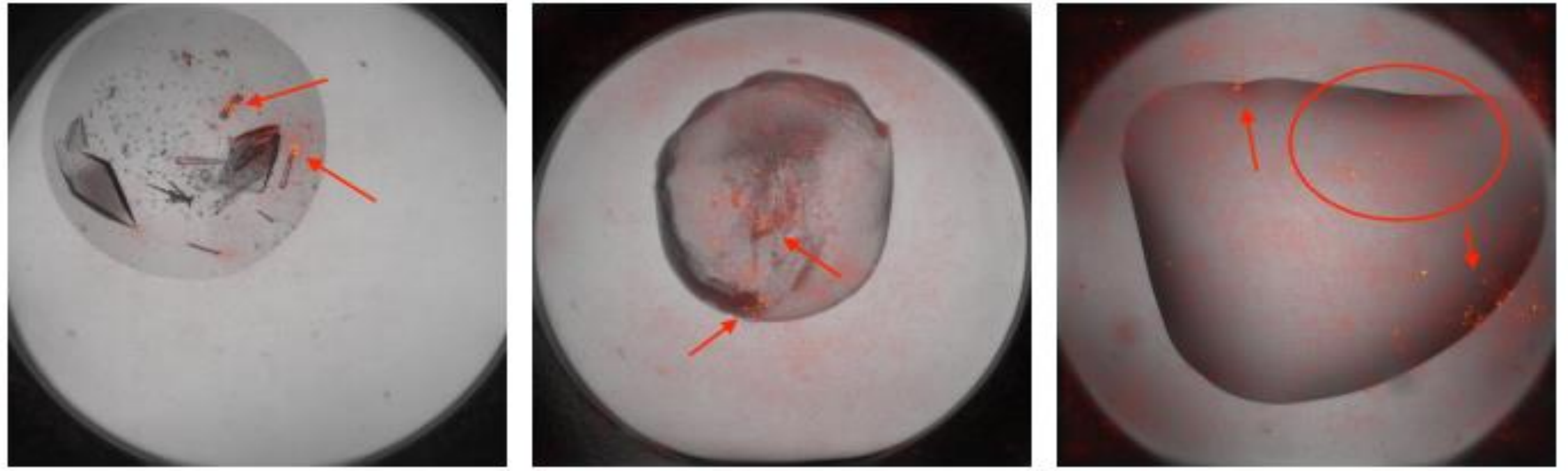
Automating Drug Discovery

- Using a rather generic CNN architecture, the authors of "Classification of crystallization outcomes using deep convolutional neural networks" were able to get 94% accuracy in identifying proteins from recrystallizations.
- This task is though because of the high variability between the imaging and data acquisition technology used.
- A single solution to work with all types of data is not trivial.

Automating Drug Discovery

- Of the 38 classes of known products, they clubbed them into 4, Crystals, Precipitate, Clear, other.
- The trained the classifier for these 4 classes, using high resolution images to account for the fine nature of the crystalline structure of proteins.
- The use of a generic classifier means they can further granulate the outputs to get back the original 38 classes

Automating Drug Discovery



(a) (b) (c)

(A) Crystal: the classifier focuses on some of the angular geometric features of individual crystals (arrows). (B) Precipitate: the classifier lands on the precipitate (arrows). (C) Clear: The classifier broadly samples the image, likely because this label is characterized by the absence of structures rather than their presence

Image Processing

Capturing and Fixing

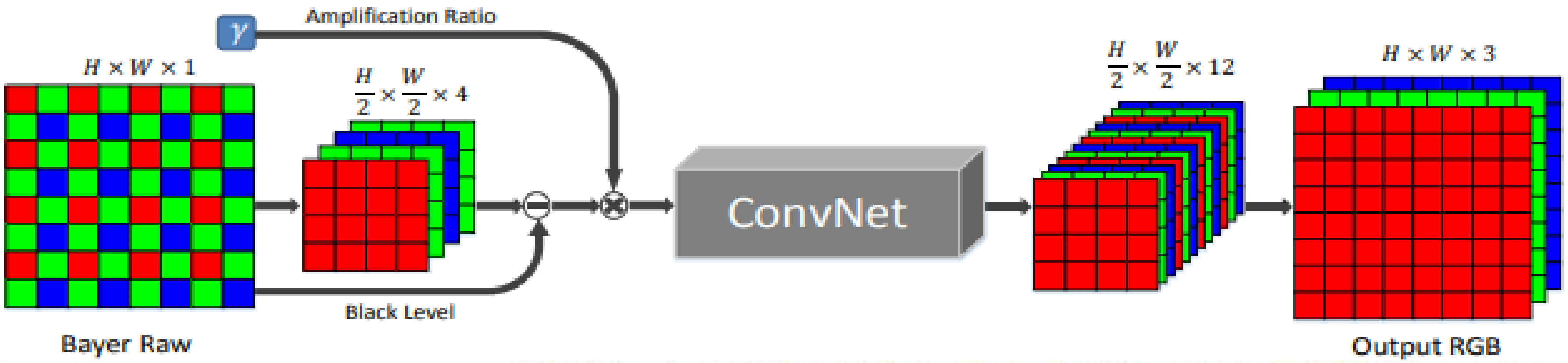
Learning To See In The Dark

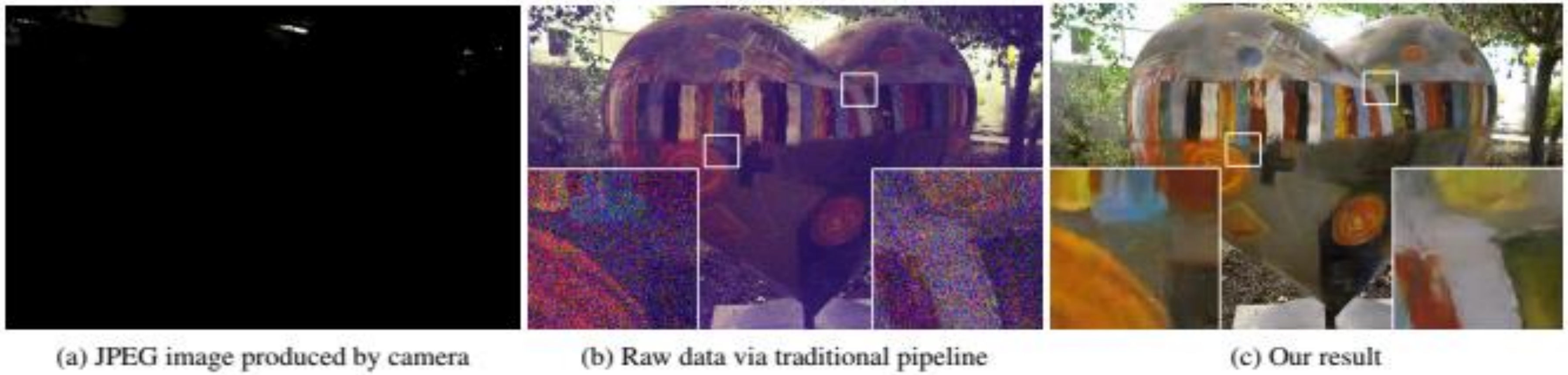
- In their latest flagship, Google introduced a feature called Night Sight, a new feature of the Pixel Camera app that lets you take sharp, clean photographs in very low light, even in light so dim you can't see much with your own eyes.
- Low light photography is hard, anybody who has photographed a dimly lit scene will be familiar with image noise, which looks like random variations in brightness from pixel to pixel.
- The exact technology used inside the camera may not come out for a while but in the paper "Learning to See in the Dark" the authors introduced a method that uses the raw output from cameras and can generate images with ample lighting from images which to the untrained eye might look pitch black

Learning To See In The Dark

- Instead of using long exposure images or bursts of images, like other traditional methods for low light photography, the authors proposed using an end-to-end CNN for handling these images.
- They used the raw output from the camera sensor, the Bayer Arrays, packed the input into four channels and correspondingly reduce the spatial resolution by a factor of two in each dimension
- They then subtract the black level and scale the data by the desired amplification ratio
- The packed and amplified data is fed into a fully-convolutional network. The output is a 12-channel image with half the spatial resolution. This half-sized output is processed by a sub-pixel layer to recover the original resolution

Learning To See In The Dark





An image captured at night by the Fujifilm X-T2 camera with ISO 800, aperture f/7.1, and exposure of 1/30 second. The illuminance at the camera is approximately 1 lux. (b) Processing the raw data by a traditional pipeline does not effectively handle the noise and color bias in the data. (c) Our result obtained from the same raw data

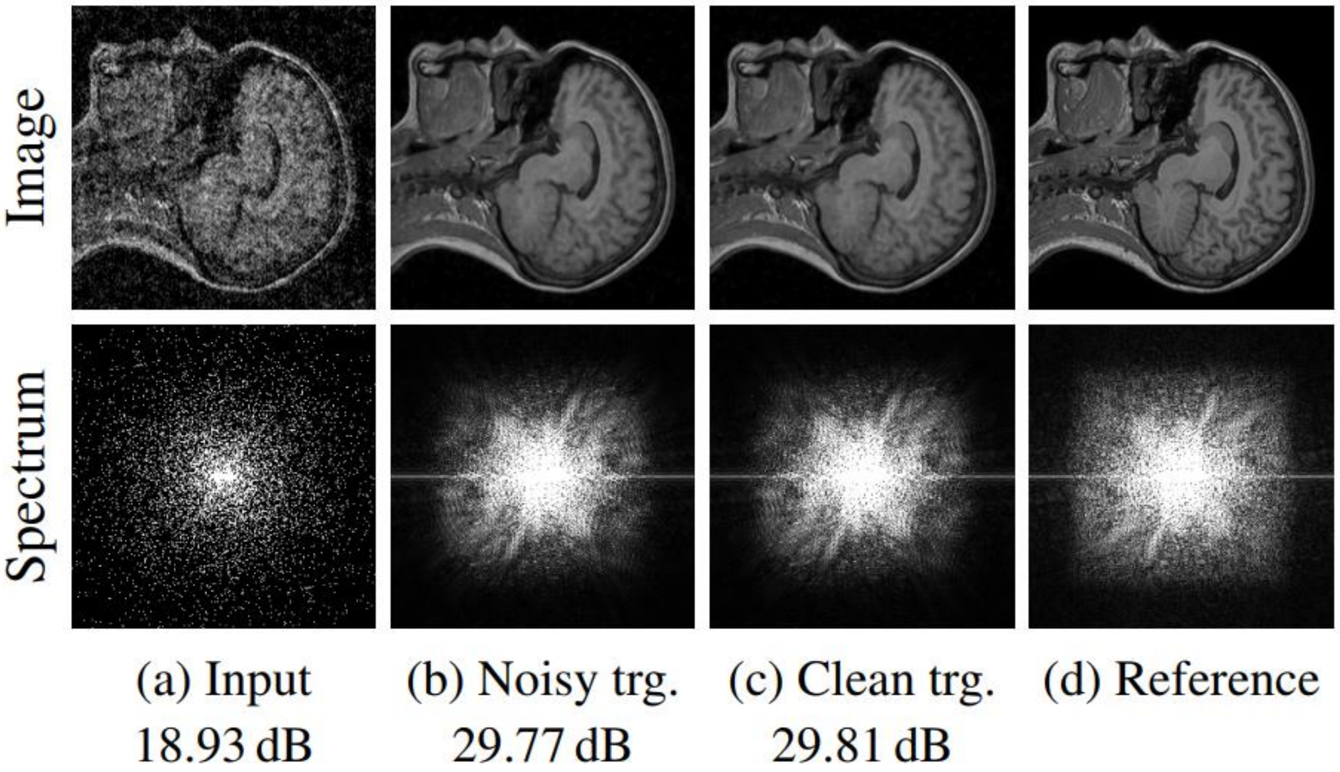
Helping MRI technicians and Animators

- Even though there’s basically nothing in common between the two professions mentioned above, both deal with the same problem: Noisy Results.
- MRI imaging process produces images with a lot of noise. Similarly, Animators use rendering methods to generate final images which are riddled with noise.
- The Authors of “Noise2Noise: Learning Image Restoration without Clean Data” proposed a novel method to removing noise from images by only seeing noisy images.
- The results have already trickled down into some image processing softwares

Helping MRI technicians and Animators

- They use an encoder-decoder style CNN for learning the noise representation. Compared to other proposed methods, their method uses a very simple model thus allowing very fast inference.

- By learning to represent noise, they're also able to perform watermark and other noise artifact removal from images.

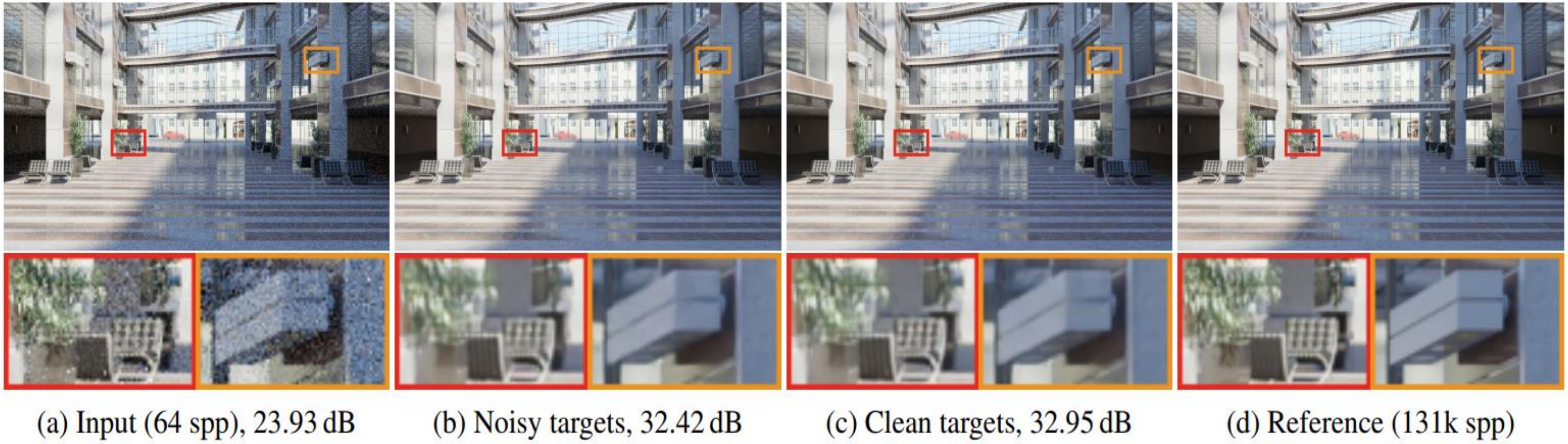


MRI reconstruction example. (a) Input image with only 10% of spectrum samples retained and scaled by 1/p. (b) Reconstruction by a network trained with noisy target images similar to the input image. (c) Same as previous, but training done with clean target images similar to the reference image. (d) Original, uncorrupted image. PSNR values refer to the images shown here, see text for averages over the entire validation set

Helping MRI technicians and Animators

- Physically accurate renderings of virtual environments are most often generated through a process known as Monte Carlo path tracing.
- This amounts to drawing random sequences of scattering events ("light paths") in the scene that connect light sources and virtual sensors, and integrating the radiance carried by them over all possible paths.
- To get clean images the simulations must run for hours if not days.
- The denoising method proposed cuts the rendering times drastically allowing for faster processing of images and bringing down costs.

Helping MRI technicians and Animators



Denoising a Monte Carlo rendered image. (a) Image rendered with 64 samples per pixel. (b) Denoised 64 spp input, trained using 64 spp targets. (c) Same as previous but trained on clean targets. (d) Reference image rendered with 131 072 samples per pixel. PSNR values refer to the images shown here, see text for averages over the entire validation set.

Recovering the Missing

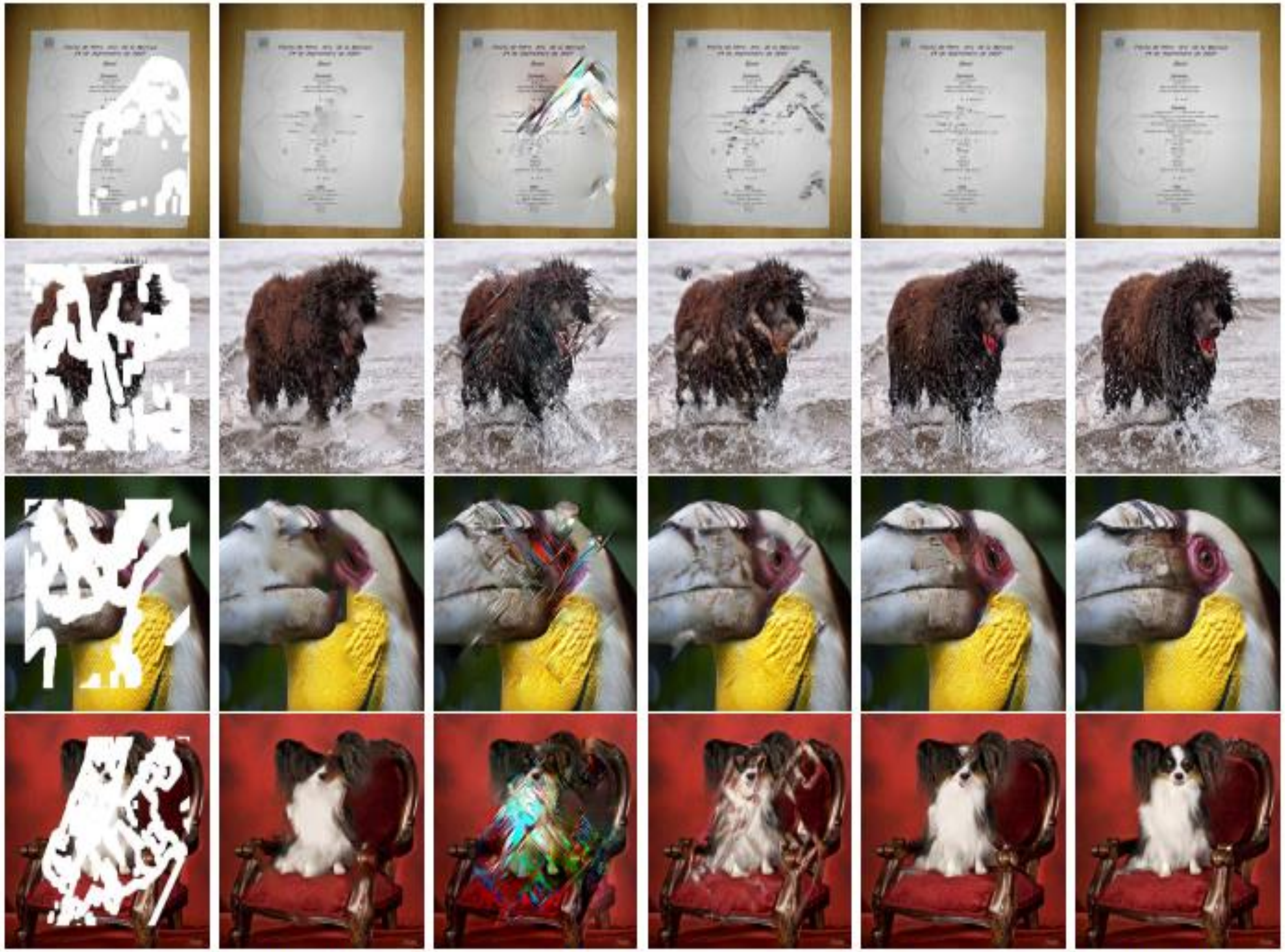
- Restoring old images is a tedious and a very involved process taking a professional days to complete.
- This include the task of filling in missing patches, removing unwanted objects and upscaling.
- Early in 2018, Guilin et. al. introduced a method doing exactly this. Following up a line of work laid by researcher at Adobe.
- They used partial convolutions, where the convolution is masked and renormalized to be conditioned on only valid pixels in their CNN model.
- Adobe's Photoshop software has shipped with previous works in the field for some time now and is soon going to absorb this work into it as well.

Recovering the Missing

- The proposed model uses stacked partial convolution operations and mask updating steps to perform image inpainting
- A partial convolution only updates the values of the output if the region on which it is processing contains pixels, i.e. the section is not missing or has been marked for removal.
- They also use a U-Net like architecture for their model. But replace all convolutional layers with partial ones and perform up sampling for the decoder
- traditional methods approximated the missing pixels from the surrounding ones, the proposed model learns the local reference by building an intuition of the regions. This also allows the model to handle cases with huge cut-outs in images.

Recovering the Missing

- Even though it is not designed for it, the model can also perform super-resolution i.e. upscaling images digitally while maintaining a high level of detail.
- To achieve this the authors, introduce holes in the image by off-setting images pixels by one and then performing in filling operations.
- They repeated this process progressively until the desired resolution is achieved.

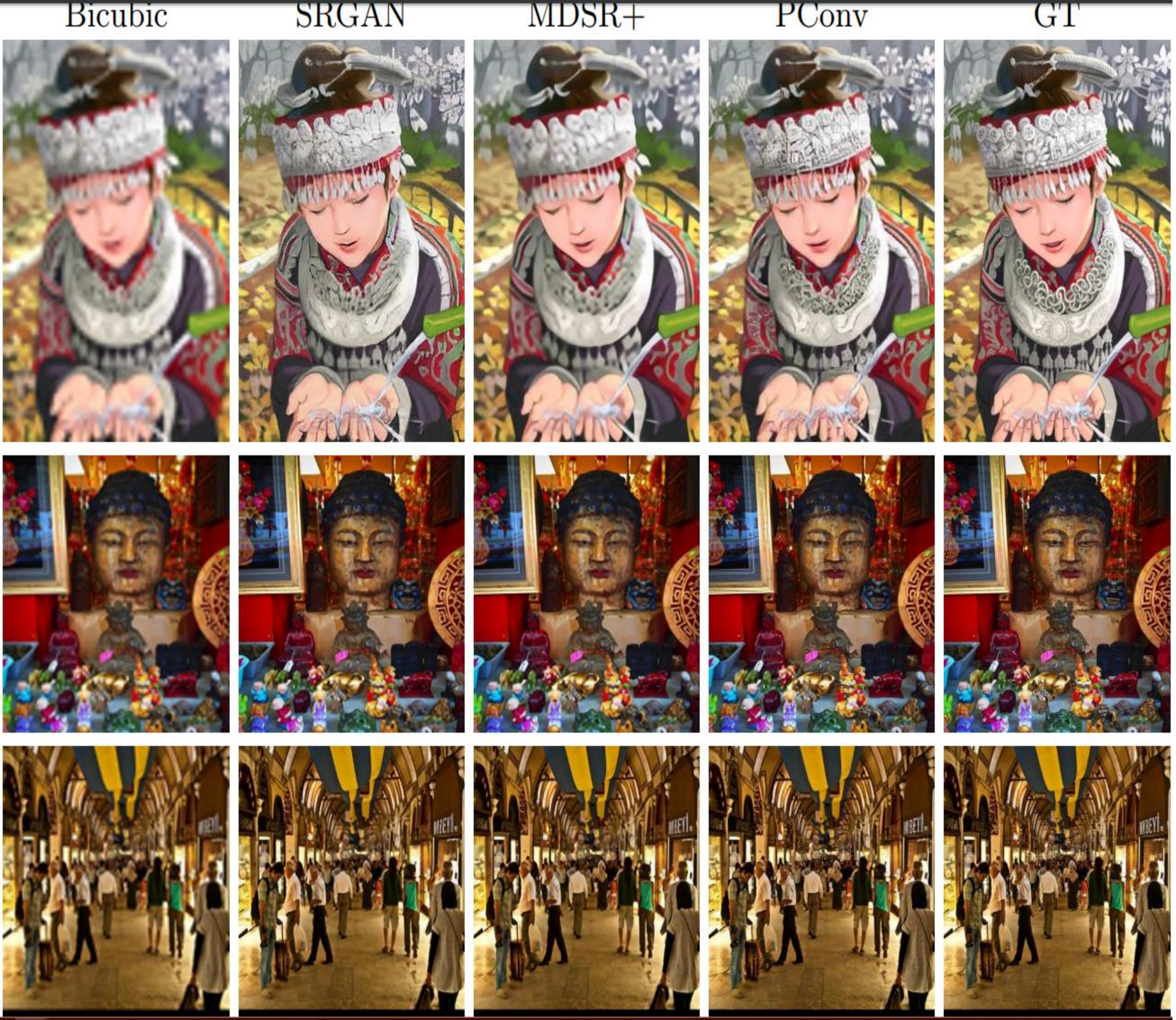


(a) Input (b) PM (c) GL (d) GntIpt (e) PConv (f) GT

Recovering the Missing

Test results on the standard ImageNet dataset for in-filing

It is clearly visible that the proposed method is far superior to existing methods.



Recovering the Missing

Results of super-resolution.
While the results are not as good the state of art it is not that far off, thus showing the generality of such model and why it has already been absorbed into commercially available products.

End

For references, definitions and further reading refer the accompanying report.