Week-12 DLP

L1 Intro. to Image Denoising & MPRNet

- some common image pre-processing steps —
  1. denoising
  2. super-resolution (increasing resolu^n by 4x)
  3. deblurring
- sources of noise
  1. photon in way of shot
  2. read noise
  3. quantiza^n noise
- image forma^n model

ground truth image + additive noise = noisy image

we want to achieve this.

- MRPNet ⇒ Multi-Stage Progressive Image Restora^n.

1. learn the contextualized features using encoder-decoder & later combining them using cross-connections (high

resolution branches) retaining the sparse info.

2. the architecture refines image quality across several stages.
3. parallel branches to extract & combine both local & global information.
4. several stage allows to move to next stage for further enhancement.

- Encoder-decoder
  1. u net type arch.
  2. Channel attenⁿ blocks (CABs) for enhanced feature extracⁿ.
  3. bilinear upsampling + conv. layers to increase spatial info.

- CAB
  1. it exploits inter-channel relationship of features.

- ORSNet - Original Resolution Subnetwork
  1. preserves fine details in image restoraⁿ.
  2. operates w/o downsampling, producing spatially enriched high-resoluⁿ features
  3. has multiple ORBs, each containing CABs.

- CSFF module → cross stage feature fusion
  1. present btw. encoder-decoder blocks.
  2. features are refined using 1 x 1 conv.
  3. reduce info. loss, enhance feature enrichment, stabilizing the network optimization process.
- SAM → supervised attenⁿ module.
  1. improves performance in multi-stage image restoraⁿ networks.
  2. provides ground-truth supervisory signals.
  3. generates attenⁿ maps to less info. features.
- loss funcⁿ.

$$L = \sum_{s=1}^{3} \left[L_{char}(X_s, Y) + \lambda \, L_{edge}(X_s, Y) \right]$$

$$L_{char} = \sqrt{\|X_s - X_Y\|^2 + \varepsilon^2}$$

$$L_{edge} = \sqrt{\|\Delta(X_s) - \Delta(Y)\|^2 + \varepsilon^2}$$

L2 Super Resoluⁿ Intro

- given a ground truth image, the task is to increase the resolution of the image. Eg. 4X.

- HR	a	b	c	d
LR - 1	$\frac{a+b}{2}$		$\frac{c+d}{2}$	

- so, according to the image forma^n model,

$$\boxed{LR = D_k H_k F_k \cdot HR}$$

scene $\Rightarrow$ HR
geometric transforma^n $\Rightarrow$ F_k
optical blur $\Rightarrow$ H_k
sampling $\Rightarrow$ D_k
Noise $\Rightarrow$ LR

- the goals of single image super-resolu^n
  1. be faithful to the low resolution input image.
  2. produce a detailed, realistic output image.

- HR $\rightarrow$ blur + downsampling $\rightarrow$ $\oplus$ (Noise) $\rightarrow$ LR

try to recover HR from LR (SISR)

∴, image super resolu^n is sometimes is ill-posed. Many solutions are possible. Now, DL comes to rescue.

- Challenges → recovering fine-texture details during large upscaling is very difficult (even in deeper CNNs)
- Problems (Current Methods) – current methods try to minimize MSE, leading to very high PSNR but lacks high freq. details & perceptual quality.
- GAN based approach → it drives the reconstruction towards the natural image manifold producing even more convincing solu^n.

L3 SRGAN

$$z \to \boxed{G} \to x \to \boxed{D} \to y = D(G(z))$$

G tries to reach 1

D tries to reach 0.

Data $\quad x \to \boxed{D} \to y = D(x)$

D tries to output 1.

- SRGAN (super resoluⁿ using Generative Adversarial network)
  ↳ It is the first framework capable of generating photo-realistic image with 4x upscale factors.
  ↳ Perceptual loss funcⁿ → combines adversarial loss and content loss
- Adversarial Network Setup -
  1. have a G and D to solve an adversarial min-max problem where G tries to fool D into classifying generative images as real.
  2. D always tries to distinguish generated images from the real ones. to encourage perceptually superior soluⁿ.
- Generator
  1. deep network with B residual blocks with 3x3 conv. layers & parametric ReLu activaⁿ.
  2. resolution is increased using Pixel shuffle layers for high quality super-resoluⁿ.

- upsampling through pixel shuffle.

$$(*, C\times r^2, H, W) \Rightarrow (*, C, H\times r, W\times r)$$

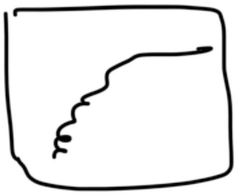
- Discriminator
  1. 8 conv. layers, with 3 x 3 filters, 2 dense layers & a sigmoid activation for binary classification.
  2. LeakyReLU ($\alpha = 0.2$) to maintain gradient flow.
  3. replaces max-pooling with strided convolutions.

L4 Image Deblurring & LAKDNet

- sources of blur — 1. camera motion
  2. object motion
  3. defocus

- point spread function or blur kernel

 ⇒ blurring

 ⇒ defocus

- A pure CNN model (LAKDnet) restores sharp, high-resolution images from blurry versions.

- Utilizes depth-wise conv. with large kernels to maintain efficiency while modeling long-range pixel dependencies.

Arch → U-shaped

Levels → 4, N LaKD blocks ⇒ N_1, N_2, N_3, N_4

Input → I (H × W × 3) and extracts low-level features.

2 main sub-modules

1. Feature Mixer – depth wise separable convolu^n with large kernel sizes (9×9). Also, has point-wise conv. layers.
   It saves a lot of computational resources.
2. Feature Fusion – has depth-wise conv. layers with 3×3 kernels local info. encoding. It has a gating mechanism with GELU activa^n to propagate & fuse features effectively.