Progress Report

Bardia Mojra

September 4, 2020

Robotic Vision Lab
The University of Texas at Arlington

1 Progress

Following items are listed in order of priority:

- Per Nolan's recommendation, I read paper [**?**] on single image multi-scale super resolution (MSSR) using CNN's. See Literature Review section for more details.

- Nolan and I are working on a report for PVBS challenge, it is due Friday. We are almost done, it should be finished by end of day, Friday.

- Have begun the dissection process for each paper I read, this is done by meticulously reading each paper and following leads.

- Began using gVim, I am still in early stages. I am ramping the use of it.

- Tried to setup Ubuntu on my machine, ran into some issues booting from USB drive. Seems to be an ASUS thing. Still working on it.

- (On pause) I am still working through Jason Brownlee's ML Mastery book, [**?**]. I am still on day 3 but need to get back to as soon as possible.

2 Plans

Following items are listed in order of priority:

- (Top priority)I will begin dissecting [**?**] to prepare for March 27th presentation.

- (On pause) Resume Robotic Perception course as soon as possible.

- (On pause) Resume Machine Learning course with Andrew Ng as soon as possible.

- (On pause) Need to read [**?**], [**?**], [**?**], and [**?**]; these papers seem fundamental to understanding the overall picture.

- (On pause) Get intimate with Python, Numpy, Pandas, Scipy, and Matplotlib, TensorFlow and PyTorch.

- (On pause) Read Digital Image Processing by Gonzalez and Woods.

- (On pause) Learn ROS.

- (On pause) (Supremely important) Read on scene understanding, semantic SLAM, graph SLAM, visual odometer, place recognition, and Kalman Filtering. Read Niko Sunderhauf's research publications.

3   General Notes

This section summarizes general research leads. The following item are to be investigated, understood and briefly summarized.

3.1   Keywords

- Bayesian Learning: This probabilistic ML approach treats model parameters as random variables. Read [?] for more details.

- Convex Optimization:

- Q-Learning:

4   Literature Review

4.1   Single Image Super-Resolution Using Multi-Scale Convolutional Neural Network - MSSR [?]

Paper proposes an architecture with two parallel path with different depths (which correspond to scales) for residual learning; where one path (module L) is used for large factor up-scaling (x4, x8) and the other (module S) for small factor up-scaling (x2). At the end, it combines the outputs by summation (a form of ensembles). In contrast to previous work where the focus is on small factor up-scaling (x2) and repeat if needed, this model takes higher factor up-scaling into consideration while training the network which helps with reducing blurriness of output image for higher factor up-scaling. The model uses multi-scale residual learning to train on general model for multiple up-scaling factors; hence, saving memory and processing time. This paper provides experiment results that show higher output image integrity where peak signal to noise ratio (PSNR) and structural similarity index (SSIM) are higher or comparatively close to state-of-the-art methods.

4.1.1 Keywords

- Bicubic interpolation: A method for artificially up-sampling an image, it is superior to nearest neighbor and bi-linear interpolation. Rather than simply averaging neighboring pixels to calculate pixel value in between, it also takes into account the gradient of its 16 neighboring pixels (hence bicubic) and results in a smoother curved image. This happens to work better than previous methods because images of physical world are made up of mostly round edges rather than sharp and square ones.

- Lanczos re-sampling:

- Statistical priors:

- Stochastic Neighbor Embedding (paper by G. Hinton [**?**]): Read paper.

- Parse coding:

- GoogLeNet:

- YCbCr color space:

- Caffe package (paper by Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.): Convolutional architecture for fast feature embedding. Read and write literature review.

- Adam method (paper by Kingma, D., Ba, J.): A method for stochastic optimization.

- PSNR (performance metric): Peak Signal to Noise Ratio represents the ratio between max (peak) possible value (power) of a signal (image) and the power of distorting noise.

- SSIM (performance metric): Structural Similarity Index, is the ratio of structural features of a processed image to the original image. The value represent percentage of structural features/information retained throughout image processing.

- A+ (SR method, paper):

- SelfEx (SR method, paper):

- SRCNN (SR method, paper): Image Super Resolution Using Deep Convolutional Networks. Read again and write literature review.

- FSRCNN (SR method, paper):

- VDSR (SR method, paper):

4.2 Value Iteration Networks - VIN [?]

4.2.1 Keywords

- Imitation Learning: **need to finish reading [?].**

- CNNs applied to reinforcement learning:

- MDP (Markov Decision Process):

- VI Algorithm (Value Iteration):

- SGD (stochastic gradient decent):

- Theano Code: "Theano is a Python library that allows you to define, optimize, and evaluate mathematical expressions involving multi-dimensional arrays efficiently. It can use GPUs and perform efficient symbolic differentiation."

- Grid-World Domain:

-

- PoseCNN, [?], uses a convolutional neural network architecture to estimate 6D pose of every object using instance segmentation with RGB-depth camera. First it uses semantic labeling to detect and classify each object within the RGB image, then it estimates the 3D translation by localizing 2D object center in the image and then estimates its distance to the camera. That is done at pixel level and then they use Hough voting method to estimate the center of the object, in 2D. This is a novel approach, but perhaps it could be improved if we first estimate the 3D contour of the object then estimate the center point of the object directly in 3D. Moreover, they use a regression model to estimate each object's 3D rotation. To train their 3D rotation regression model, authors propose two L2 loss functions, PoseLoss (PLoss) and ShapeMatch-Loss (SLoss) where SLoss does not require the specification of object symmetries. At last, they use depth image/video for improving 6D pose annotation and training their model in a global optimization step. Similar learning models can be developed using Deep

Reinforcement Learning since ground truth is available via depth camera.