Machine learning model research

# Neuroevolution (unplanned)

Can be used when a score is provided rather than a incorrect or correct value score. Apparently this can be applied to the LSTM

# LSTM (unimplemented)

LSTM’s are a type of reccurent neural networks, however avoid the gradient vanishing problem with the aid of internal

# LSTM Alternatives

## Modular Memory Units (unplanned)

A [paper](http://jenjenchung.github.io/anthropomorphic/Papers/Khadka2018neuroevolution.pdf) describes the benefits of MMU’s due to the external position of the memory compared to LSTM’s memory being held within the cell.

# Global inputs

Include inputs that provide information that is constant throughout the PCB. This would include board size, board shape, number of layers.

# Loss Function

## Mean Rounding Loss (unimplemented)

A dense NN provides floating point values but the layer variables need to be integers

# Mixture Density Networks

Add in modelling the inverse problems as they tend to have multiple solutions. Regular dense networks assume the conditional probability distributions is gaussian while MDN use multiple components to construct more complex distributions. When components are gaussian for continuous and Bernoulli for binary. The least squares method assumes a gaussian distribution. Parameters often seen: mixture coeffeictients (pi), means (mu) and varian (sigma squared). Takes x as input and outputs theta components (usually gaussian) to create a more complicated distribution. MDN combine a feed forward (right figure) network with a mixture model (left figure). Number of outputs (L+2)K. error function: maximum likelihood

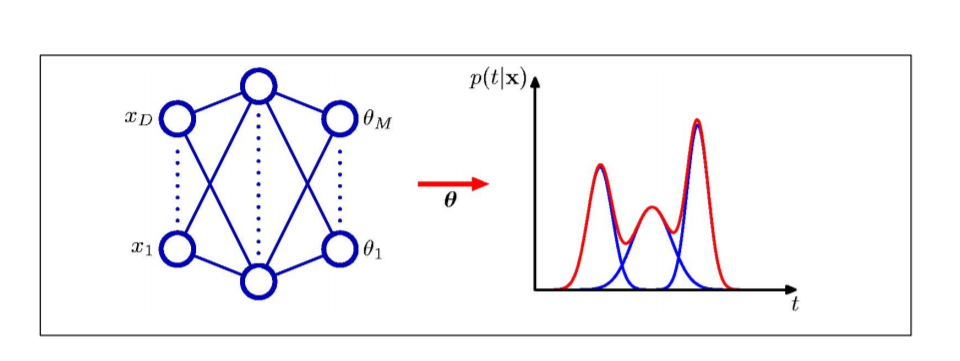
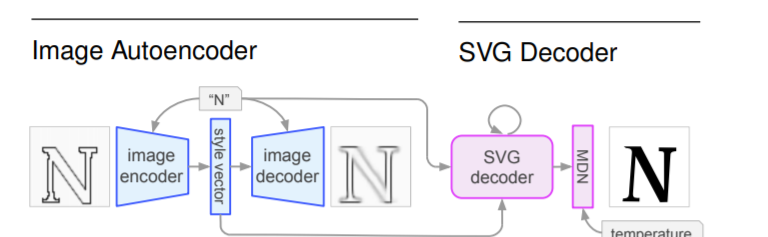


Figure 1 https://cedar.buffalo.edu/~srihari/CSE574/Chap5/Chap5.6-MixDensityNetworks.pdf

# Works read

## A learned representation for scalable graphics

* Data set size of 14M
* Other works use a GAN for image generation. Alternatively probalistic models with invertible representations have also been seen.
* Autoregressive models employ sequational prediction
  + Commonly use teacher-forcing
* Non differentiable setups use REINFORCE algorithm for adversarial training
* 
* Simarlity is quantified with a convolutional variational autoencoder (vae): blue
* Instead of a dense convolutional network after the decoder (LSTM) they use a mixture density network (MDN)
* Tensor2Tensor library used
* Z representation of raster images
* Used relative path positioning for svg creation. Along with a rescaling
* Trained in two parts encoder-decoder then svg decoder (blue and pink)
* VAE and MDN are probalistic so 10 samples were taken with best performer picked
* Uses a log likelihood for MDM (pink). Cross entropy loss of command type and MDN loss applied to real valued arguments
* Uses adam optimiser
* Suggested improvements attention based architectures and adversarial training. Additionally the ability for the model to error correct

### Questions raised

* Can I use a variational auto encoder or something like generate the latent representation? The training of this compression would then be independent of pcb track routing stage.
* Should I use a MDN or a Dense