Evolving Deep Neural Networks

Gabriel Meyer-Lee, Harsha Uppili, Alan Zhuolun Zhao

December 4, 2017

Agenda

1. Goal and Motivation
2. Background
3. Implementation
    i. CNN
    ii. LSTM
4. Experiment
    a. Image Classification
    b. Language Modeling
        i. Text Classification
        ii. Text Generation
5. Results
    a. Image
    b. Text
6. Conclusion
7. Future Work

Goals

- **Goals**
- Background
- Implementation
- Experiment
- Results
- Conclusion
- Future Work

Goal

“The brain—**including its architecture and how it learns**—is a product of natural evolution” - Kenneth O. Stanley

...So why not do the same for DNNs?

*Successfully develop an **evolutionary scheme to evolve deep neural networks**

*Apply this scheme to **image classification and language modeling tasks**

Background

- Goal
- **Background**
- Implementation
- Experiment
- Results
- Conclusion
- Future Work

Optimizing Deep Neural Networks

- Bayesian Optimization (Snoek et. al. 2015)
- Deep Reinforcement Learning (Zoph et. al. 2017)
- Large Scale Evolution (Real et. al. 2017)
- CoDeepNEAT (Mikkulainen et. al. 2017)

- Goal
- **Background**
- Implementation
- Experiment
- Results
- Conclusion
- Future Work

Hierarchical Representations

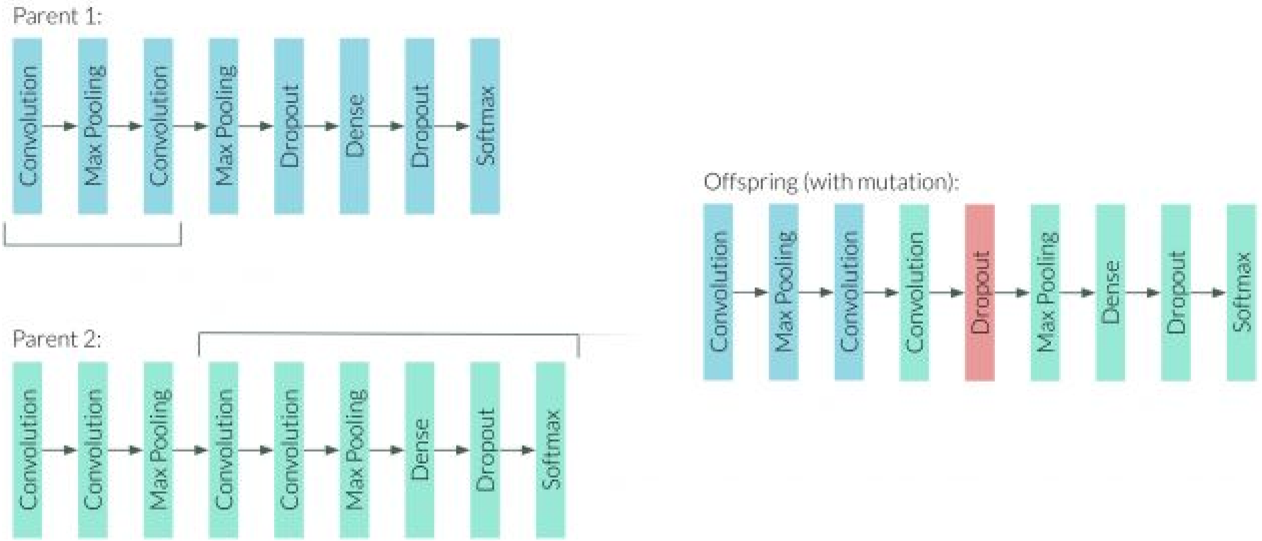
- HSANE (Moriarty, Mikkulainen 1998)
  - Evolves at neuron level
- CoDeepNEAT (Mikkulainen et. al. 2017)
  - Two co-evolved hierarchical levels
- Hierarchical Representations for Efficient Architecture Search
  - Based on Real et. al. (2017)'s graph based deep neural network representation

Implementation

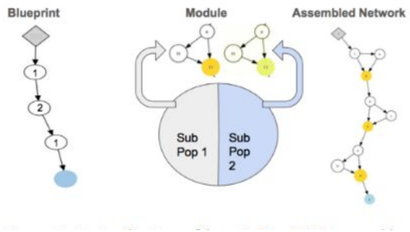
- Goal
- Background
- **Implementation**
- Experiment
- Results
- Conclusion
- Future Work

DEvol

- First attempt
- Fixed-length genome encoding



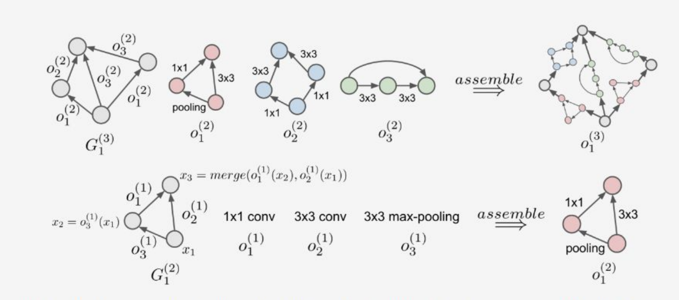
CoDeepNEAT



- Based as closely as possible on the details given in the preprint from Mikkulainen et. al.
- Evolves deep networks with a stacked architecture. Co-evolves the height of the stack and modules included (Blueprints chromosomes) and the make up of the modules (Module chromosomes)
- Capable of evolving convolutional, dense, and LSTM layers
- Differences from paper:
  - Input corruption and global hyperparameters
  - LSTM implementation

- Goal
- Background
- **Implementation**
- Experiment
- Results
- Conclusion
- Future Work

Hierarchical



- Based on a ICLR 2018 submission, “Hierarchical Representations for efficient architecture search”
- Uses simple single population evolution and random search
- Represents deep neural networks as self-similar graphs. At the base level each node is feature map and each edge is a layer (here only convolutional or pooling layers). At higher levels each node is still a feature map but each edge is a member of the previous level.
- Fixed node sizes and motif sizes for each level, still allows complexification through identity and no - op connections.
- Evolves cells for fixed blueprint architecture.

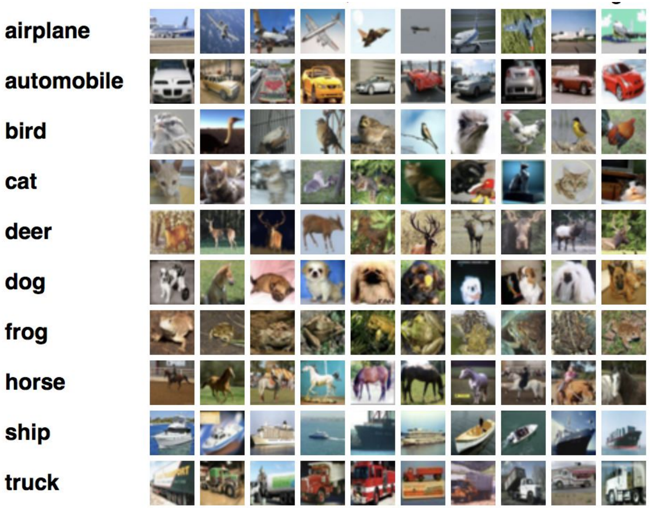
Experiments - Image Classification

- Goal
- Background
- Implementation
- **Experiment**
- Results
- Conclusion
- Future Work

Image Classification

- Cifar 10
  - Dataset of 32x32 color images, labeled over 10 categories
  - 50000 training images, 10000 testing images

- Comparison with MNIST
  - Color images
  - Real world objects
  - Large variance



- Goal
- Background
- Implementation
- **Experiment**
- Results
- Conclusion
- Future Work

DEvol

- Training setup
  - EvoTraining set: 42500
  - Validation set: 7500
  - Max convolutional layers: 6
  - Max dense layers : 3
  - 50 Generation
  - Population size: 50
- Training Procedure
  - Image Pre-processing
  - Run evolution with training on EvoTraining set for 10 epochs
  - Fitness function: performance on validation set
  - Train best performers with whole training set (50000) and test the accuracy on test set (10000)

- Goal
- Background
- Implementation
- **Experiment**
- Results
- Conclusion
- Future Work

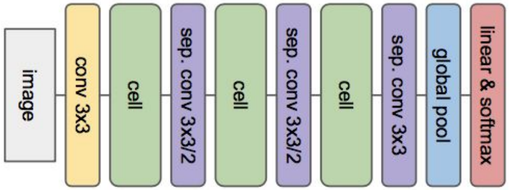
CoDeepNEAT

- Training Setup
  - EvoTraining set: 42500
  - Validation set: 7500
  - 25 Generation
  - Population size:
    - 15 modules
    - 10 blueprints
    - 25 models generated with blueprints and modules
- Training Procedure
  - Image Pre-processing
  - Run evolution with training on EvoTraining set for 8 epochs
  - Fitness function: performance on validation set
  - Train best performers with whole training set (50000) and test the accuracy on test set (10000)

- Goal
- Background
- Implementation
- **Experiment**
- Results
- Conclusion
- Future Work

Hierarchical Representation

- Training Setup
  - EvoTraining set: 42500
  - Validation set: 7500
  - Pre-defined blueprint
  - Population size:
    - Start with 10 modules (models)
  - 100 steps (random/evolution)
- Training Procedure
  - Image Pre-processing
  - Run steps with training on EvoTraining set for 10 epochs
  - Fitness function: performance on validation set
  - Train best performers with whole training set (50000) and test the accuracy on test set (10000)



small CIFAR-10 model

Experiments - Language Modeling

- Goal
- Background
- Implementation
- **Experiment**
- Results
- Conclusion
- Future Work

Experiment - Text Classification

```
<REUTERS TOPICS=''YES'' LEWISSPLIT=''TRAIN''
CGISPLIT=''TRAINING-SET'' OLDID=''12981'' NEWID=''798''>
<DATE> 2-MAR-1987 16:51:43.42</DATE>
<TOPICS><D>livestock</D><D>hog</D></TOPICS>
<TITLE>AMERICAN PORK CONGRESS KICKS OFF TOMORROW</TITLE>
<DATELINE> CHICAGO, March 2 - </DATELINE><BODY>The American Pork
Congress kicks off tomorrow, March 3, in Indianapolis with 160
of the nations pork producers from 44 member states determining
industry positions on a number of issues, according to the
National Pork Producers Council, NPPC.
Delegates to the three day Congress will be considering 26
resolutions concerning various issues, including the future
direction of farm policy and the tax law as it applies to the
agriculture sector. The delegates will also debate whether to
endorse concepts of a national PRV (pseudorabies virus) control
and eradication program, the NPPC said. A large
trade show, in conjunction with the congress, will feature
the latest in technology in all areas of the industry, the NPPC
added. Reuter
\&\#3;</BODY></TEXT></REUTERS>
```

Class	P(C)
Acquisitions	**.43**
Gain	.21
Crude	.06

...

Article → Word2Vec → Padding → Complete

One Hot Encoding

- Goal
- Background
- Implementation
- **Experiment**
- Results
- Conclusion
- Future Work

Text Generation

```
Yields on money-market mutual funds continued to slide, amid signs that portfolio managers expect further declines in interest rates.

The average seven-day compound yield of the 400 taxable funds tracked by IBC/Donoghue's Money Fund Report eased a fraction of a percentage point to 8.45% from
8.47% for the week ended Tuesday.
Compound yields assume reinvestment of dividends and that the current yield continues for a year.

Average maturity of the funds' investments lengthened by a day to 41 days, the longest since early August, according to Donoghue's.
Longer maturities are thought to indicate declining interest rates because they permit portfolio managers to retain relatively higher rates for a longer period.
Shorter maturities are considered a sign of rising rates because portfolio managers can capture higher rates sooner.

The average maturity for funds open only to institutions, considered by some to be a stronger indicator because those managers watch the market closely, reached a
high point for the year -- 33 days.

Nevertheless, said Brenda Malizia Negus, editor of Money Fund Report, yields "may blip up again before they blip down" because of recent rises in short-term
interest rates.
The yield on six-month Treasury bills sold at Monday's auction, for example, rose to 8.04% from 7.90%.
```

n partially-overlapping sequences

n one hot encoding vectors of the next character

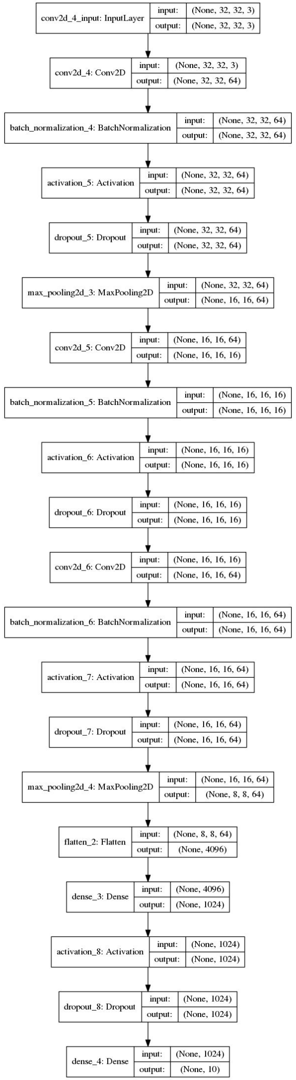
`Typically, money-fund yields` → ...

Results

DEvol

- Goal
- Background
- Implementation
- Experiment
- **Results**
- Conclusion
- Future Work

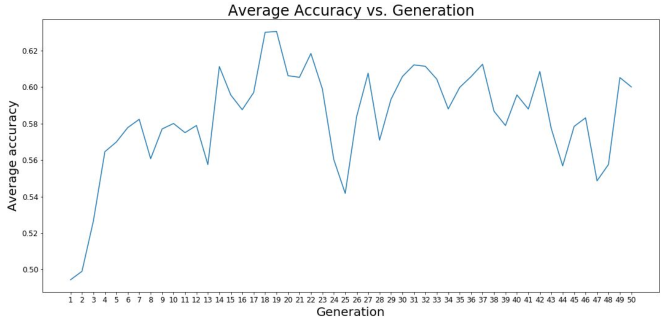
- Linear model

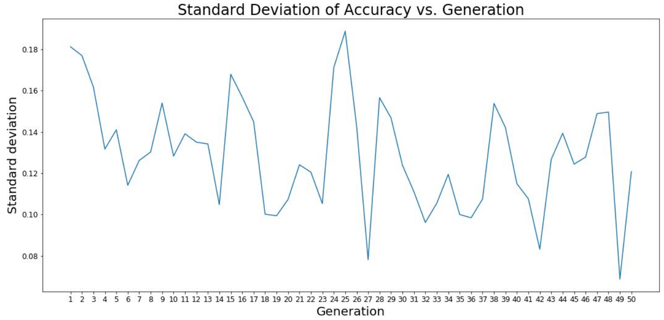


DEvol

- Goal
- Background
- Implementation
- Experiment
- **Results**
- Conclusion
- Future Work

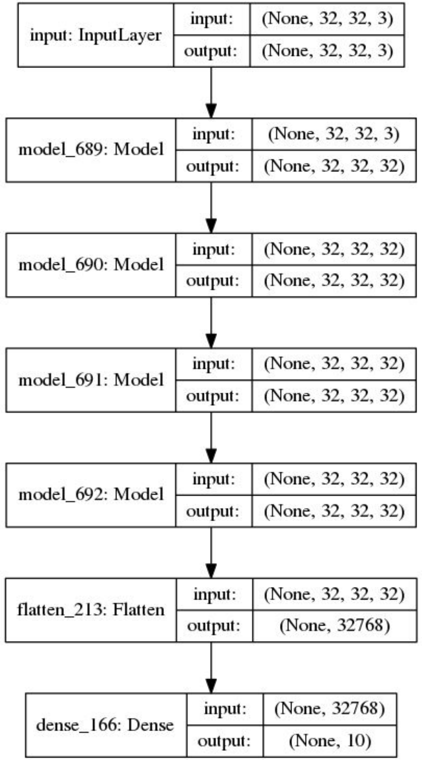
- Evolution pattern is not very significant
- Chose moderately complex structures over highly complex structures: 866 conv3 vs. 48 conv6



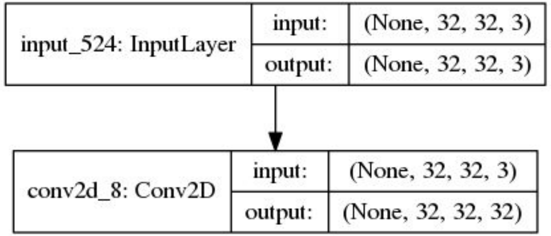


- Goal
- Background
- Implementation
- Experiment
- **Results**
- Conclusion
- Future Work

CoDeepNEAT (still running)



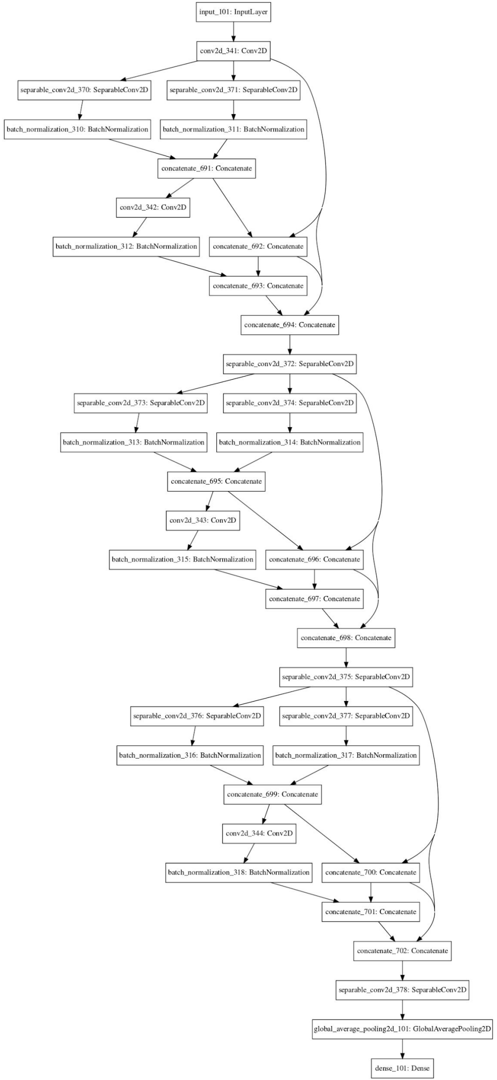
- 7/25 Generation
- Still linear



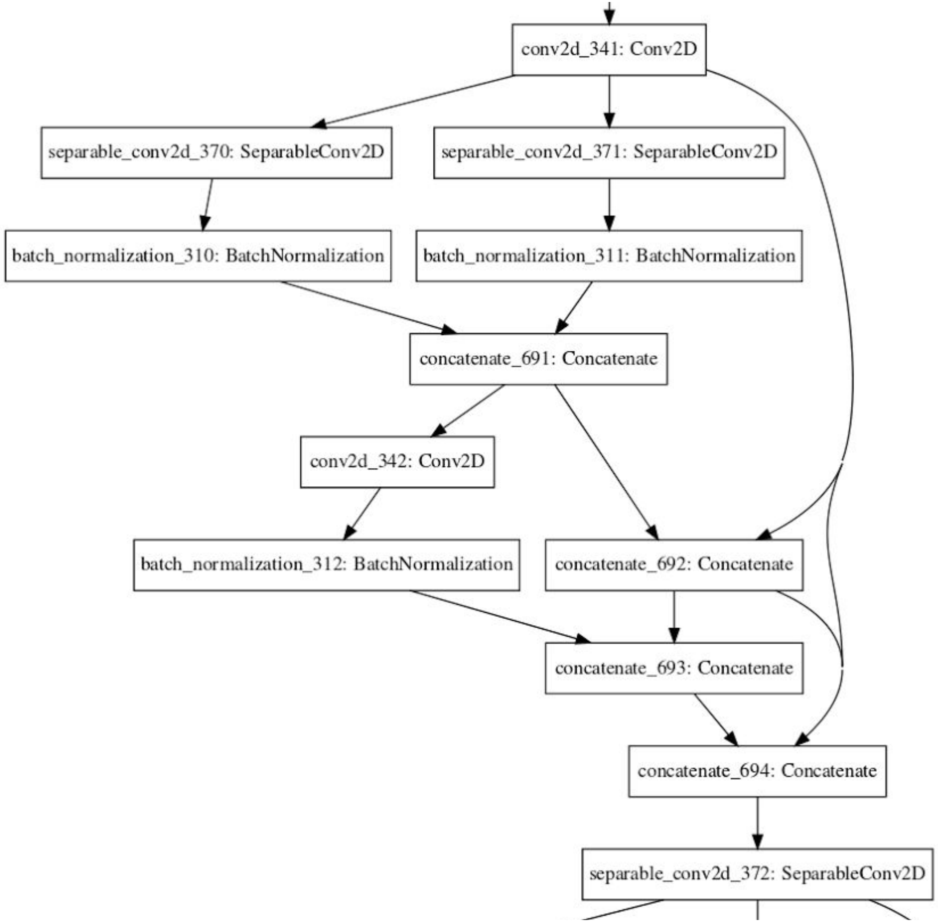
- Goal
- Background
- Implementation
- Experiment
- **Results**
- Conclusion
- Future Work

HR

- Random steps
- It's BIG (and non-linear)



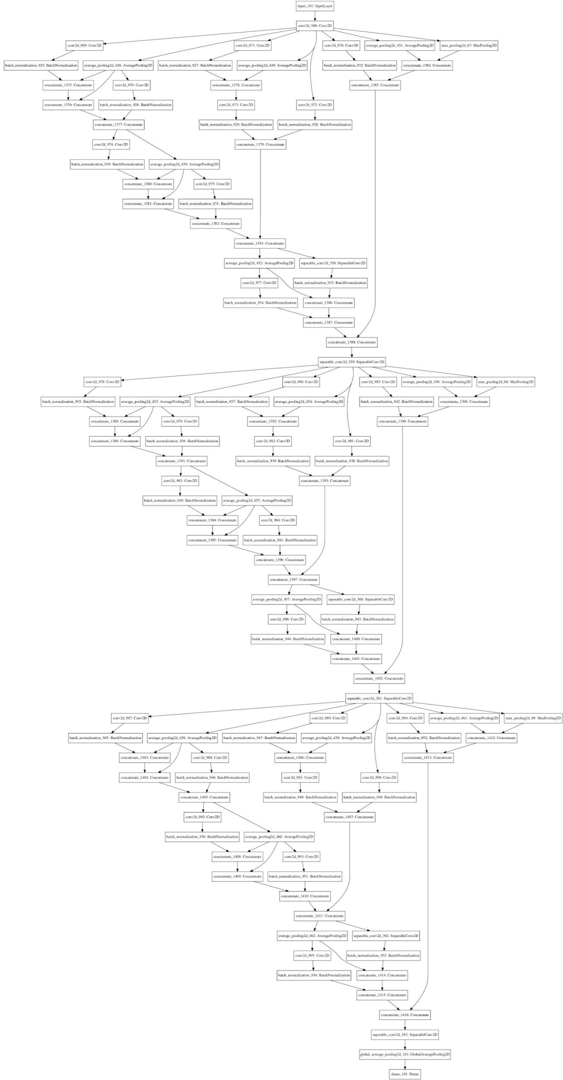
- Goal
- Background
- Implementation
- Experiment
- **Results**
- Conclusion
- Future Work



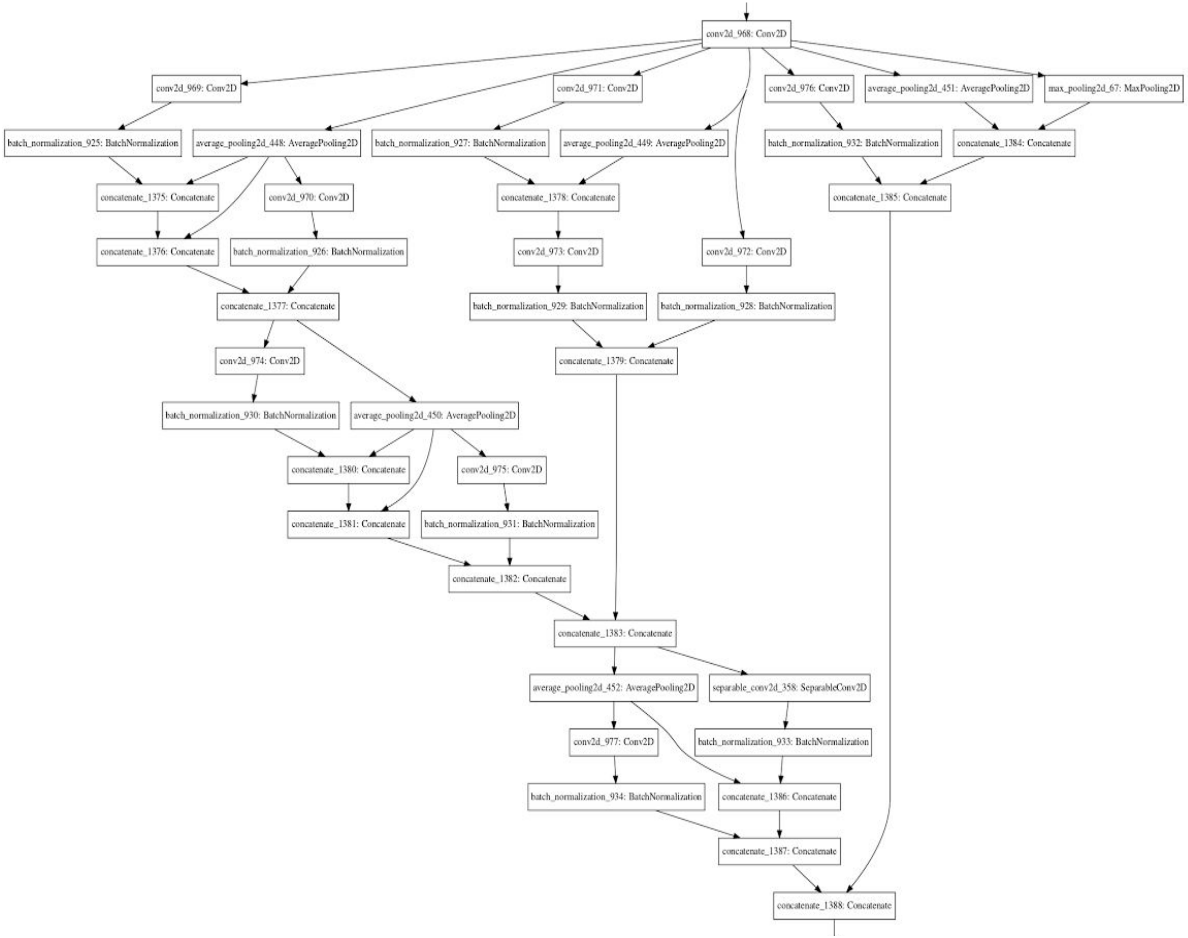
- Goal
- Background
- Implementation
- Experiment
- **Results**
- Conclusion
- Future Work

HR

- Evolution steps
- It's YUUUUUUGE



- Goal
- Background
- Implementation
- Experiment
- **Results**
- Conclusion
- Future Work



- Goal
- Background
- Implementation
- Experiment
- **Results**
- Conclusion
- Future Work

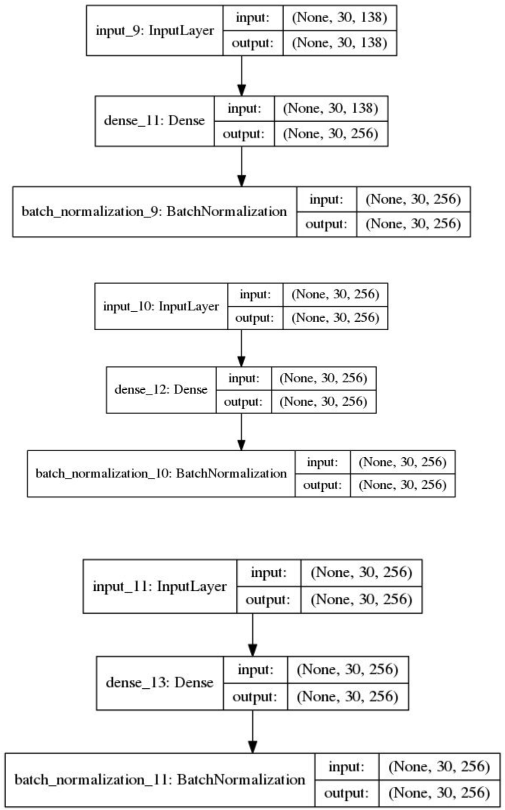
Performance Comparison

Algorithm	# GPU*days	Test Accuracy
DEvol	2.89	76.5%
CoDeepNEAT (7 Generations)	0.4	65.2%
Hierarchical	0.5	Random: 78% (100 ep.) Evolve: 74% (10 ep.)
Real et al.	2750	94.6%
Zoph et al.	1800	96.6%

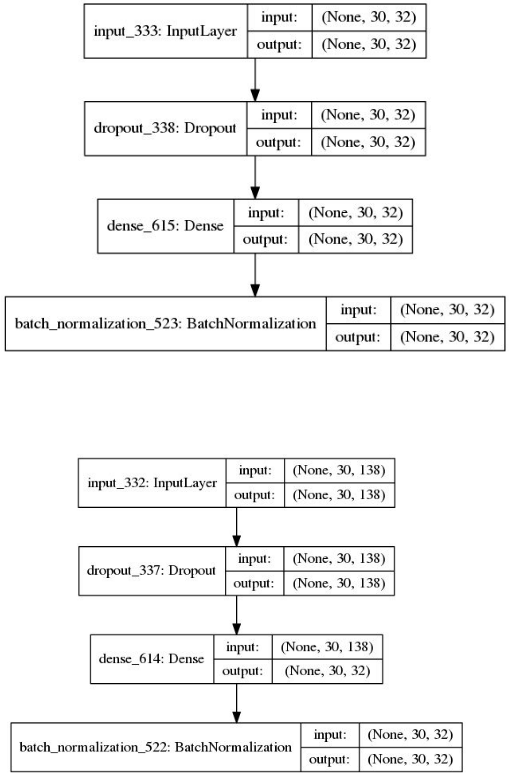
- Goal
- Background
- Implementation
- Experiment
- **Results**
- Conclusion
- Future Work

Text Classification - Modules

First Model



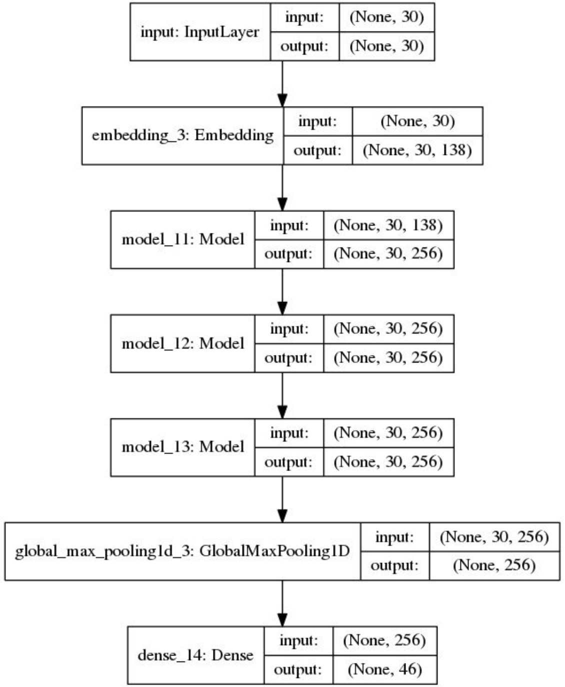
Last Model



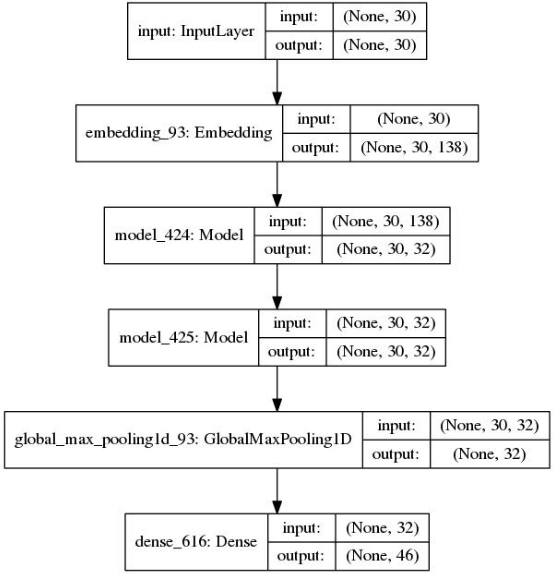
- Goal
- Background
- Implementation
- Experiment
- **Results**
- Conclusion
- Future Work

Text Classification - Models

First Model



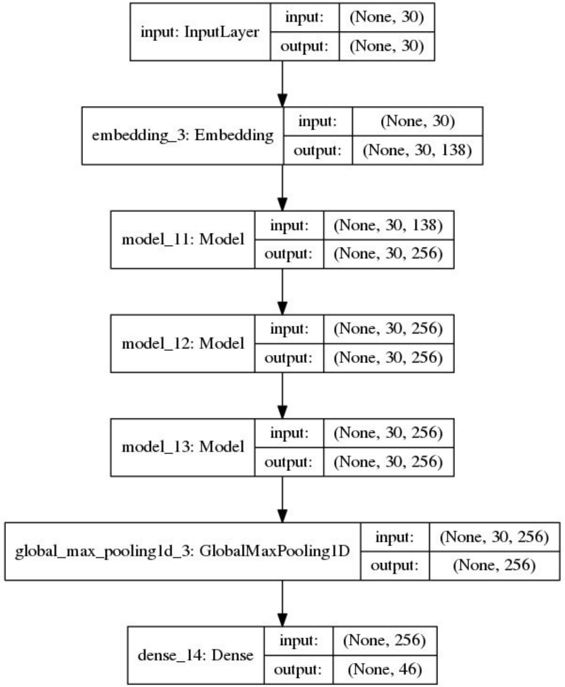
Last Model



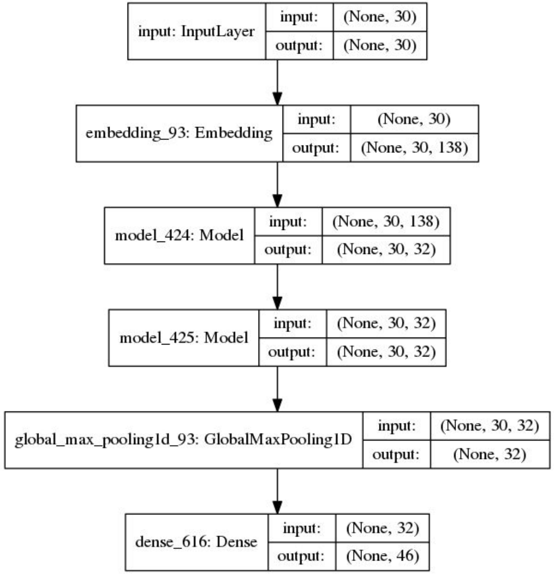
Text Classification - Models

- Goal
- Background
- Implementation
- Experiment
- **Results**
- Conclusion
- Future Work

First Model

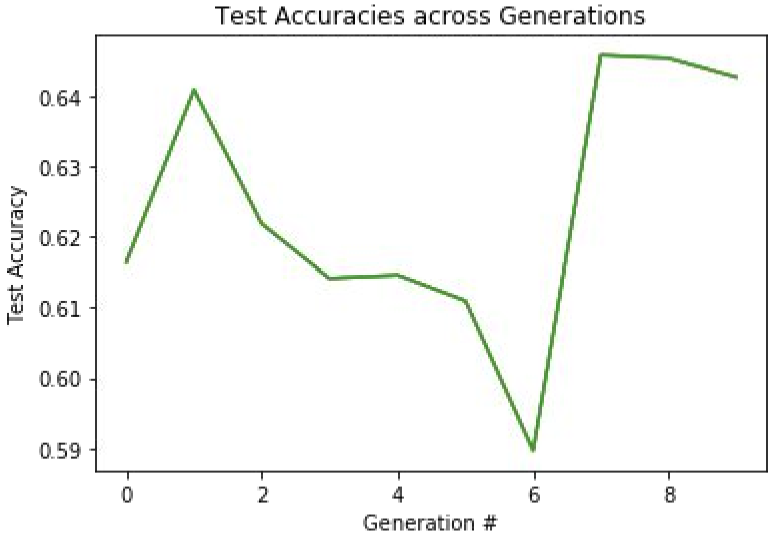


Last Model



Text Classification - Test Accuracies

- Goal
- Background
- Implementation
- Experiment
- **Results**
- Conclusion
- Future Work



Test Accuracy with single-layer
LSTM Network: 88.97%

Conclusions (so far)

- Goal
- Background
- Implementation
- Experiment
- Results
- **Conclusion**
- Future Work

- Evolving deep neural networks takes a long time
- Evolving non-linear architecture is important for deep CNN
- Hierarchical Representation was best at discovering non-linear structures, but also relatively biased
- CoDeepNEAT was the least biased but need better evolving mechanism for non-linear architecture
- CoDeepNEAT will recognize when LSTM layers are necessary

Future Work

- Goal
- Background
- Implementation
- Experiment
- Results
- Conclusion
- **Future Work**

Can we successfully perform text generation?

→ Compare with literature

Implement graph-based Hierarchical LSTM architecture search

More testing is necessary!

Thank you!
Any questions?