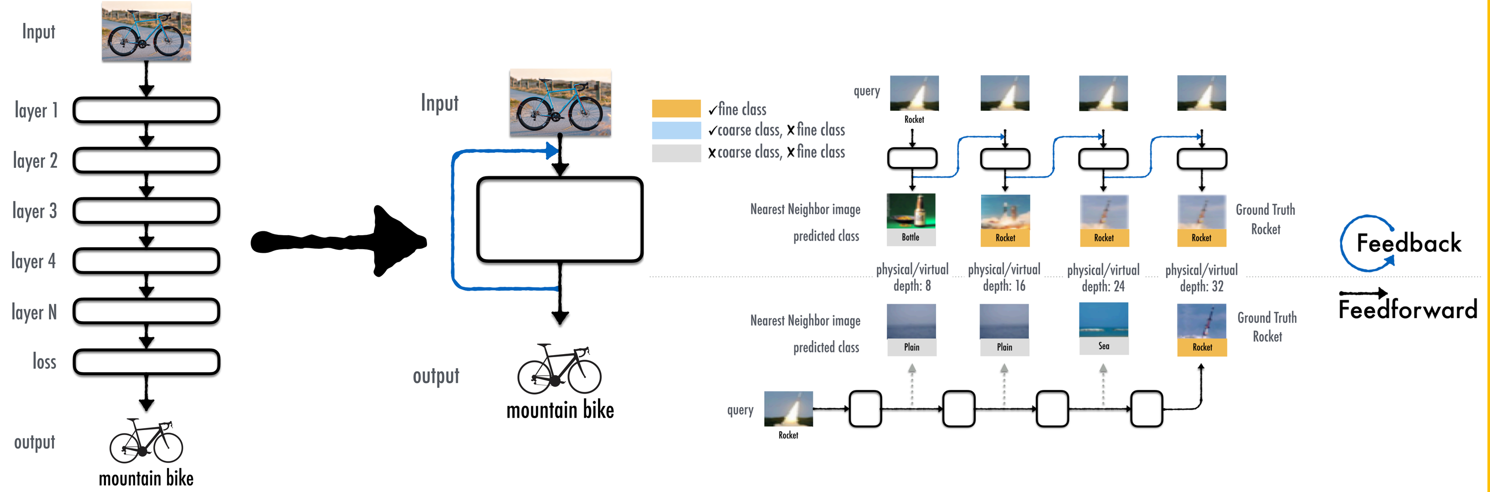
# Feedback networks

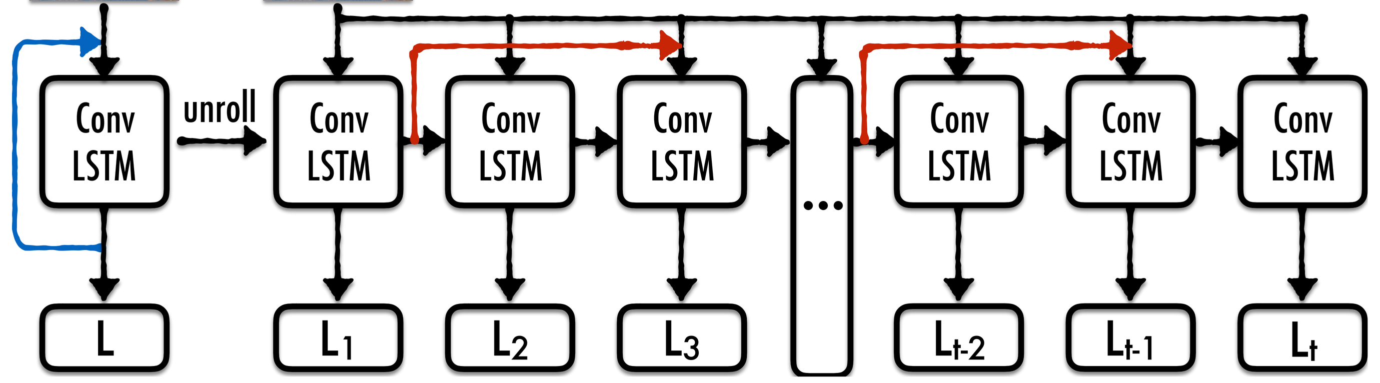
Amir R. Zamir* Te-Lin Wu* Lin Sun William B. Shen
Jitendra Malik Silvio Savarese
http://feedbacknet.stanford.edu

- **From Feedforward to Feedback**
  - Information passed via hidden states (ConvLSTM)
  - Feedback: Notion of output at each iteration
- **Advantages:**
  - Early Predictions at the Query Time
  - Taxonomic Compliance
  - New basis for Curriculum Learning
- **Representations:** predictions happen in an iterative manner coupled with a coarse-to-fine representation

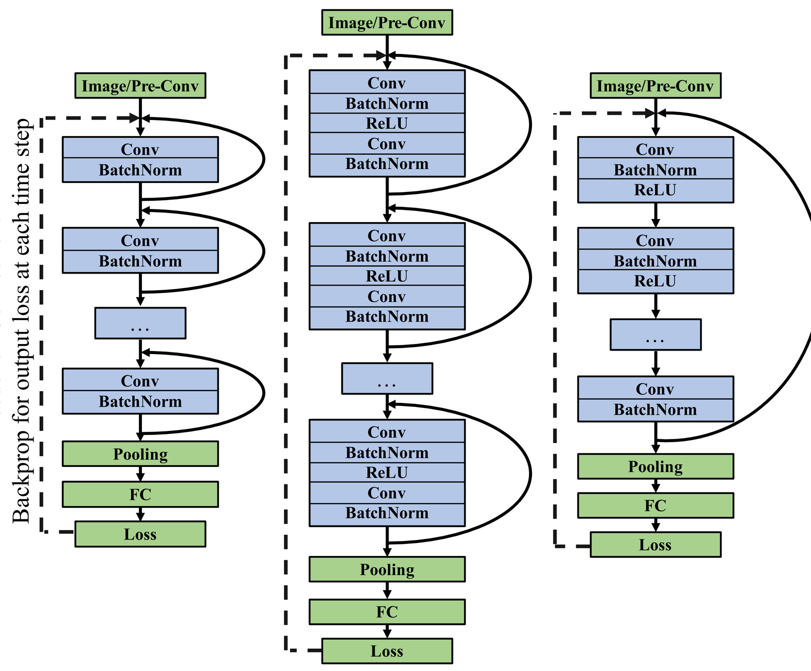



- **Core Feedback Model:**

- **Red Lines:** Temporal Skip Connection


- **Different Feedback Modules**



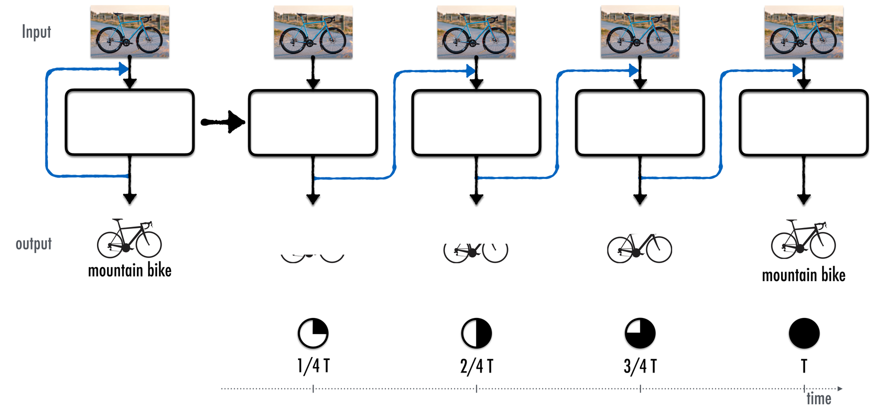

| Feedback Type | Top1 | Top5 |
|---|---|---|
| Stack-1 | 66.29 | 89.58 |
| Stack-2 | **67.83** | **90.12** |
| Stack-All | 65.85 | 89.04 |

- **Suggests the local feedback length is neither too short nor too long.**

| Feedback Net | Top1 | Top5 |
|---|---|---|
| w/o skip connections | 67.37 | 89.97 |
| w/ skip connections | **67.83** | **90.12** |

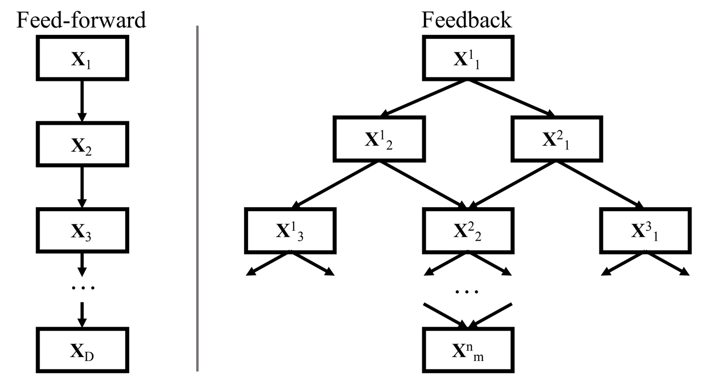
- **Comparison of with and without the temporal skip inference**

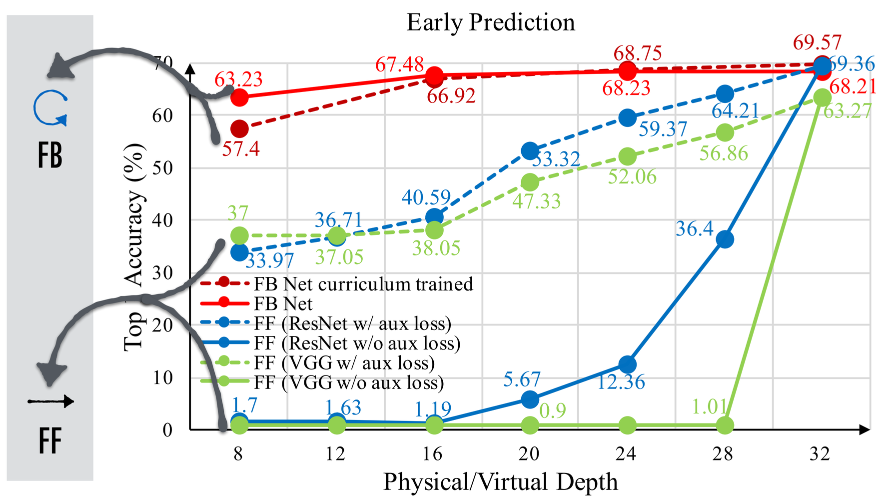
- **Advantage I: Early Prediction**




- Provides estimation of the output in a fraction of the total inference time

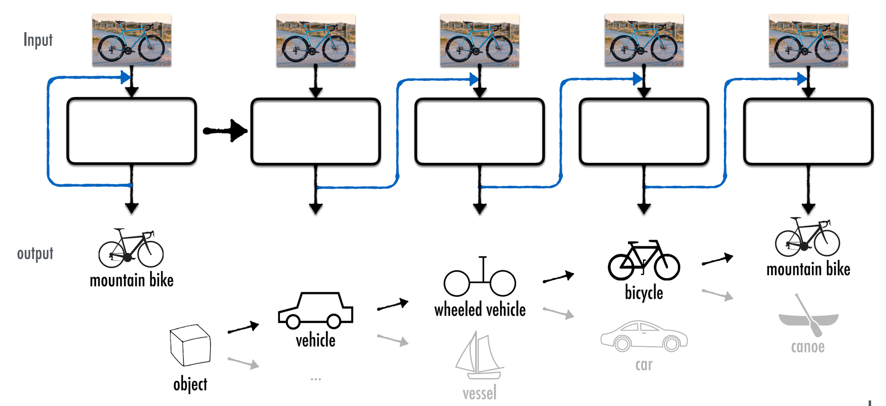
- The computational graph is also one order shallower than the feedforward



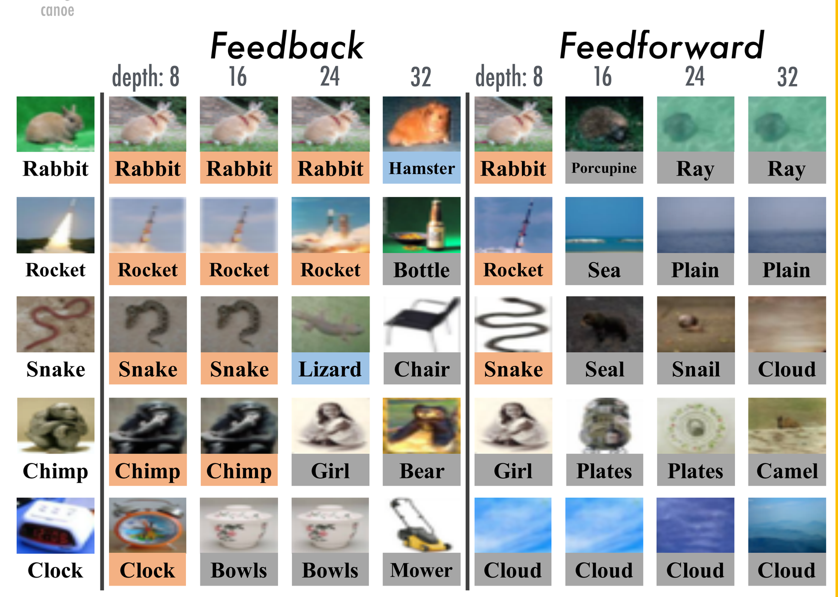

- Comparison of accuracy of feedback model and the feedforward baselines

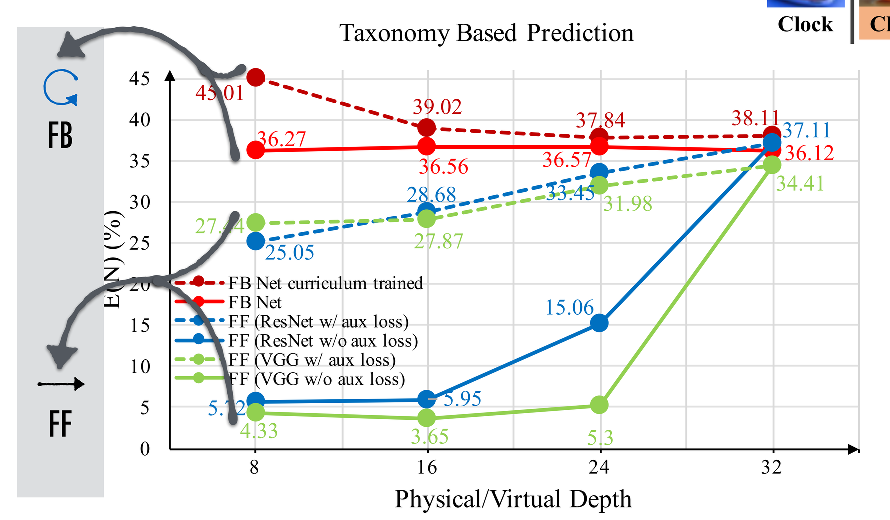
- **Advantage II: Taxonomic Compliance**




- Predictions naturally conform to a hierarchical structure in the output space (a taxonomy)

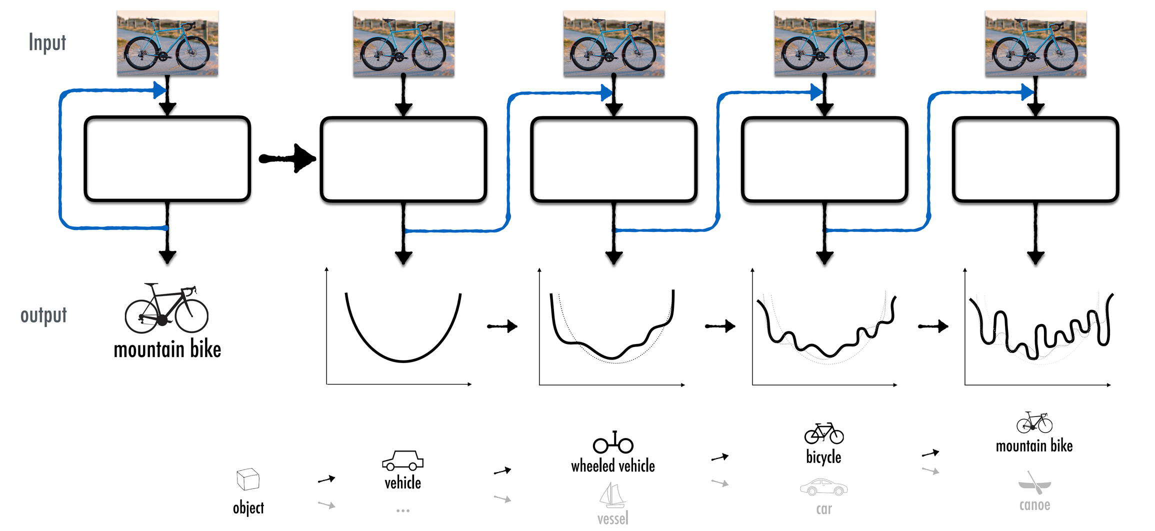
- **Qualitative Results**: Each row shows a query along with nearest neighbors at different depths for feedback and feedfoward networks




- The probability of making a correct coarse prediction for a query if it made a wrong fine prediction for it

- **Advantage III: Episodic Curriculum Learning**




| CIFAR100 | | | |
|---|---|---|---|
| Model | CL | Fine | Coarse |
| Feedback Net | N | 68.21 | 79.7 |
| | Y | **69.57**(+1.34%) | **80.81**(+1.11%) |
| Feedforward | N | 69.36 | 80.29 |
| ResNet w/ Aux loss | Y | 69.24(-0.12%) | 80.20(-0.09%) |
| Feedforward | N | 69.36 | 80.29 |
| ResNet w/o Aux loss | Y | 65.69(-3.67%) | 76.94(-3.35%) |
| Feedforward | N | 63.56 | 75.32 |
| VGG w/ Aux loss | Y | 64.62(+1.06%) | 77.18(+1.86%) |
| Feedforward | N | 63.56 | 75.32 |
| VGG w/o Aux loss | Y | 63.2(-0.36%) | 74.97(-0.35%) |

- Evaluation of the impact of Curriculum Learning (CL) on CIFAR100.
- Curriculum training also improves taxonomic prediction which it is condition on

- Feedback enables enforcing on episodic curriculum.
- Any hierarchical output space or taxonomy can be used as a curriculum strategy.
- We use annealed loss function at each time step

$$L(t) = \zeta L_t^{Coarset} + (1-\zeta) L_t^{Fine}$$

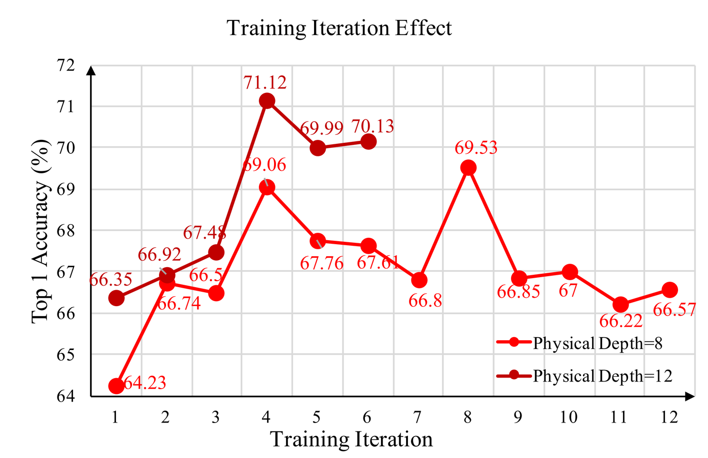
- Evaluation of the impact of Curriculum Learning (CL) on Stanford Car dataset.

| Stanford Cars Dataset | | | |
|---|---|---|---|
| Model | CL | Fine | Coarse |
| Feedback Net | N | 50.33 | 74.15 |
| | Y | **53.37**(+3.04%) | **80.7**(+6.55%) |
| Feedforward | N | 49.09 | 72.60 |
| ResNet-24 | Y | 50.86(+1.77%) | 77.25(+4.65%) |
| Feedforward | N | 41.04 | 67.65 |
| VGG-24 | Y | 41.87(+0.83%) | 70.23(+2.58%) |

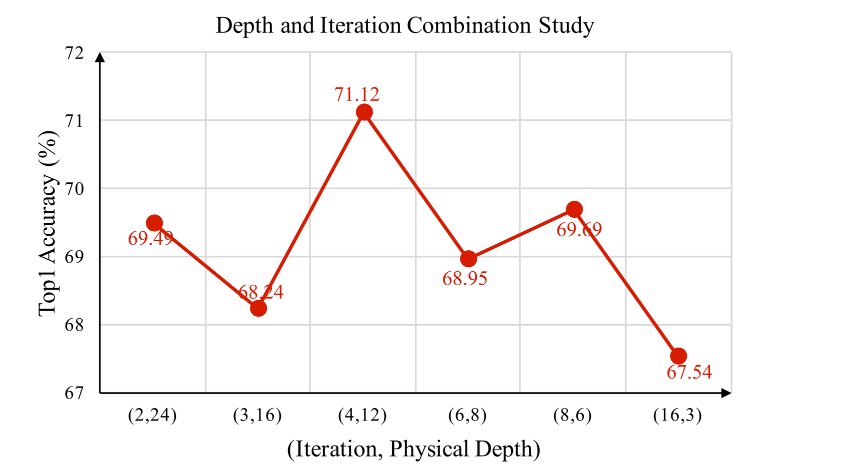
- **End-Point Performances and Analysis**

| CIFAR100 | | | | |
|---|---|---|---|---|
| Model | P/D | V/D | Top1 (%) | Top5 (%) |
| Feedforward (ResNet[13]) | 48 | - | 70.04 | 90.96 |
| | 32 | - | 69.36 | 91.07 |
| | 12 | - | 66.35 | 90.02 |
| | 8 | - | 64.23 | 88.95 |
| | 128* | - | 70.92 | 91.28 |
| | 110* | - | 72.06 | 92.12 |
| | 64* | - | 71.01 | 91.48 |
| | 48* | - | 70.56 | 91.60 |
| | 32* | - | 69.58 | 91.55 |
| Feedforward (VGG[37]) | 48 | - | 55.08 | 82.1 |
| | 32 | - | 63.56 | 88.41 |
| | 12 | - | 64.65 | 89.26 |
| | 8 | - | 63.91 | 88.90 |
| **Feedback Net** | 12 | 48 | **71.12** | **91.51** |
| | 8 | 32 | 69.57 | 91.01 |
| | 4 | 16 | 67.83 | 90.12 |

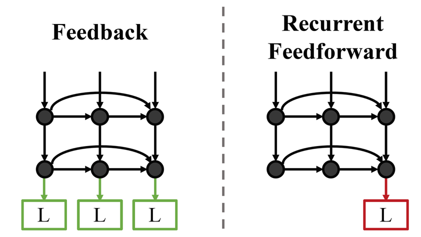
| Stanford Cars dataset | | | |
|---|---|---|---|
| Model | CL | Fine | Coarse |
| Feedback Net | N | 50.33 | 74.15 |
| | Y | **53.37**(+3.04%) | **80.7**(+6.55%) |
| Feedforward | N | 49.09 | 72.60 |
| ResNet-24 | Y | 50.86(+1.77%) | 77.25(+4.65%) |
| Feedforward | N | 41.04 | 67.65 |
| VGG-24 | Y | 41.87(+0.83%) | 70.23(+2.58%) |




- CIFAR100 performance comparison of the same physical structure trained for different iteration.



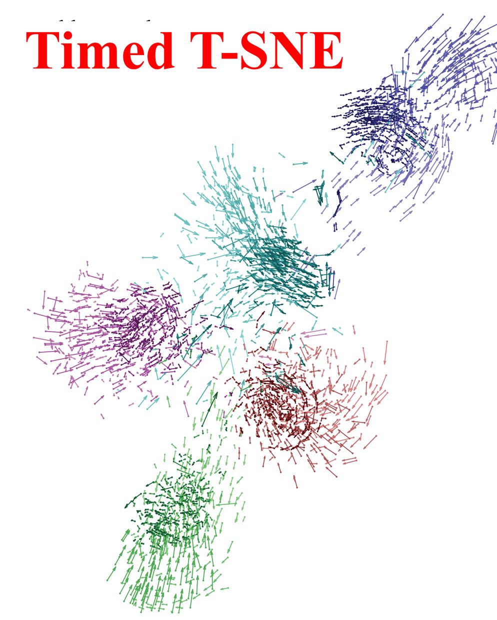

- CIFAR100 performance comparison of feedback models with same virtual depth but different (iteration, physical depth) combinations.



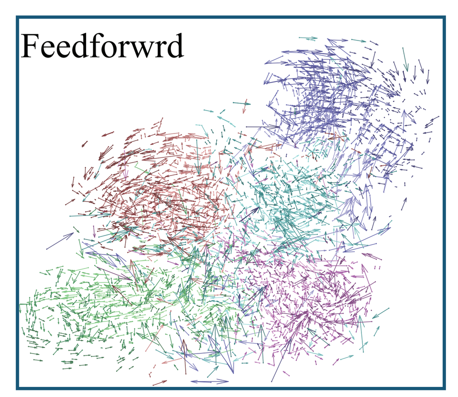

- The major difference between our model and recurrent feedforward models is how the loss is connected

| | Virtual Depth | | | |
|---|---|---|---|---|
| Model | 12 | 24 | 36 | 48 |
| Feedback | **67.94** | **70.57** | **71.09** | 71.12 |
| Feedback Disconnected (Recurrent Feedforward) | 36.23 | 62.14 | 67.99 | **71.34** |

- **Timed T-SNE**



- Illustrates how the representation of a network evolves throughout depth/iterations




- **Feedback v.s. ResNet Ensemble:** comparison between Feedback Net and an ensemble of ResNets that produce early predictions at the same computation graph depth time steps

| | Time Steps | | | |
|---|---|---|---|---|
| Model | 12T | 15T | 18T | 21T |
| Feedback Network | **67.94** | **70.57** | **71.09** | **71.12** |
| ResNet Ensemble | 66.35 | 67.52 | 67.87 | 68.2 |

*Demo, Data, Code, Results:*
*http://feedbacknet.stanford.edu*