# Feedback Networks

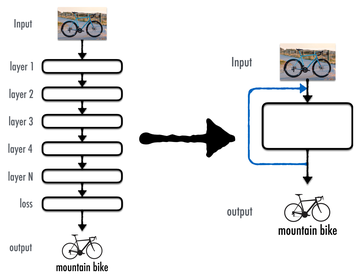
Amir R. Zamir* Te-Lin Wu* Lin Sun William B. Shen Bertram E. Shi Jitendra Malik Silvio Savarese

Stanford University — Berkeley UNIVERSITY OF CALIFORNIA

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

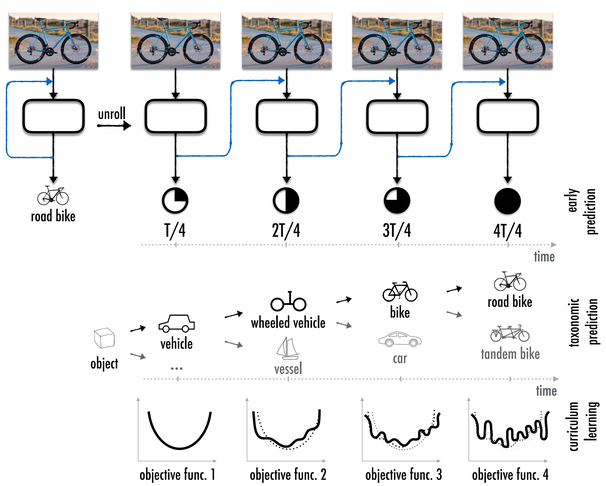
## From Feedforward to Feedback

- A general-purpose feedback based learning model
- Can replace feedforward models as a blackbox
- Can be instantiated using existing recurrent models



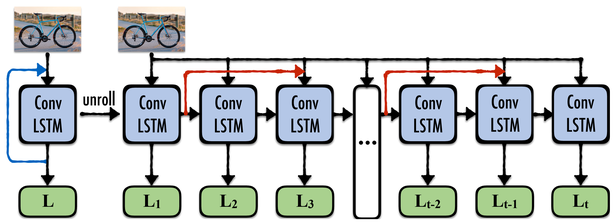

- **Core Advantages:** (see below ↓ )
  - Early predictions at query time
  - Taxonomy compliance in output
  - New basis for Curriculum Learning
- **Internal Representation:** coarse-to-fine. Meaningfully and notably different from feedforward representation.

## Core Advantages:



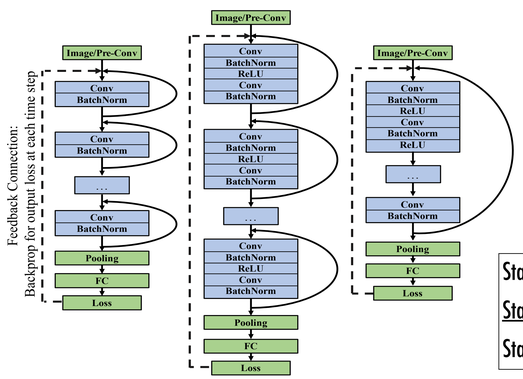

## Feedback Model Details

- **Feedback Definition**: when a system's output is routed back into input as part of an iterative cause-and-effect process [13].
- ⇒ **Two Requirements**: (1) iterativeness, (2) rerouting a notion of posterior (output) back into the system in each iteration.
- Can instantiate feedback use existing RNNs (ConvLSTM[66]).




$$\mathbf{L} = \sum_{t=1}^{T} \gamma^t \mathbf{L_t}, \text{ where } \mathbf{L_t} = -log \frac{e^{\mathbf{H_t^D}[C]}}{\sum_j e^{\mathbf{H_t^D}[j]}}.$$

## Feedback Modules and Their Lengths




| | |
|---|---|
| Stack-1: | 66.29% |
| Stack-2: | 67.83% |
| Stack-All: | 65.85% |

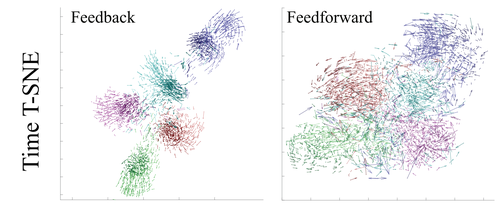
## Episodic Curriculum Learning

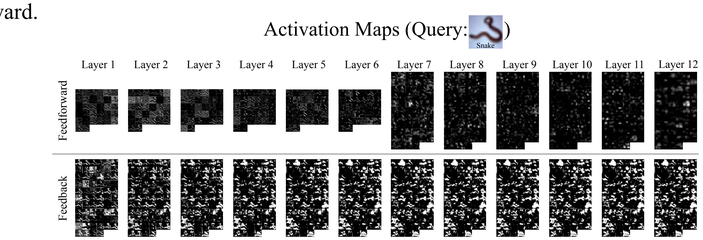
- Enabled by feedback (unlike feedforward).
- Any hierarchical output space or taxonomy can be used as a curriculum strategy.
- We use annealed loss function at each iteration.

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

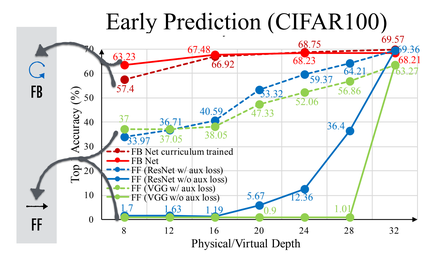
## The Internal Representation:

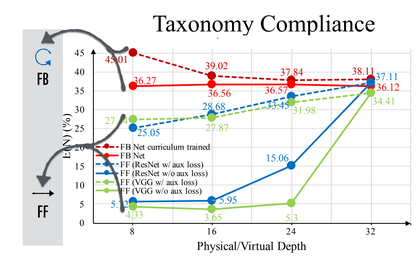
- Feedback develops a course-to-fine representation. Unlike low-abstraction to high-abstraction of feedforward.
- Feedback's representation is notably dissimilar to feedforward's.

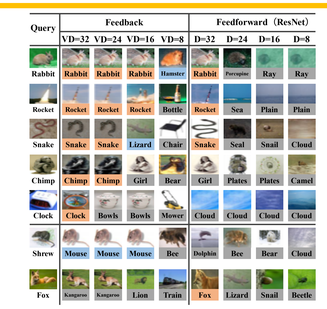



## Quantitative Results




### Endpoint Results (CIFAR100)

| Model | Physical Depth | Virtual Depth | Top1 (%) | Top5 (%) |
|---|---|---|---|---|
| **Feedback Net** | 12 | 48 | **71.12** | **91.51** |
| | 8 | 32 | 69.57 | 91.01 |
| | 4 | 16 | 67.83 | 90.12 |
| Feedforward (ResNet[19]) | 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[47]) | 48 | - | 55.08 | 82.1 |
| | 32 | - | 63.56 | 88.41 |
| | 12 | - | 64.65 | 89.26 |
| | 8 | - | 63.91 | 88.90 |

| Model | Virtual Depth 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** |

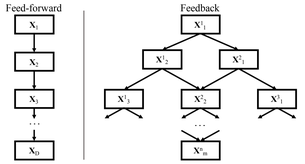
### Endpoint Results (Stanford Cars)

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

### Curriculum Learning (CIFAR100)

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

### Computation Graphs




### Depth & Iteration Combination Study