# Supplementary Material: Task-Driven Modular Networks for Zero-Shot Compositional Learning

Senthil Purushwalkam[1][*]      Maximilian Nickel[2]      Abhinav Gupta[1,2]      Marc'Aurelio Ranzato[2]

[1]Carnegie Mellon University        [2]Facebook AI Research

## 1. Hyperparameter tuning

The results we reported in the main paper were obtaining using the best hyper-parameters found on the validation set. We used the same cross-validation procedure for all methods, including ours. Here, we present the ranges of hyper-parameters used in the grid-search and the selected values.

### 1.1. Task Driven Modular Networks

Hyper-parameter values:

- Feature extractor learning rates: 0.1, 0.01, 0.001, 0.0001 (chosen: 0.001)

- Gating network learning rates: 0.1, 0.01, 0.001, 0.0001 (chosen: 0.01)

- Number of sampled sampled negatives for Eq 3: for MIT States 200, 400, 600 (chosen: 600), for UT-Zappos we choose all negatives

- Batch size: 64, 128, 256, 512 (chosen: 256)

- Fraction of train concepts dropped in ConceptDrop: 0%, 5%, 10%, 20% (chosen: 5%)

- Number of modules per layer: 12, 18, 24, 30 (chosen: 24)

- Output dimensions of each module: 8, 16 (chosen: 16)

- Number of layers: 1, 2, 3, 5 (chosen: 3 for MIT States, 2 for UT-Zappos)

### 1.2. LabelEmbed+

Hyper-parameter values:

- Learning rates: 0.1, 0.01, 0.001, 0.0001 (chosen: 0.0001 for MIT States, 0.001 for UT-Zappos)

- Batch size: 64, 128, 256, 512 (chosen: 512)

- Fraction of train concepts dropped in ConceptDrop: 0%, 5%, 10%, 20% (chosen: 5%)

### 1.3. RedWine

Hyper-parameter values:

- Learning rates: 0.1, 0.01, 0.001, 0.0001 (chosen: 0.01)

- Batch size: 64, 128, 256, 512 (chosen: 256 for MIT States, 512 for UT-Zappos)

- Fraction of train concepts dropped in ConceptDrop: 0%, 5%, 10%, 20% (chosen: 0%)

### 1.4. Attributes as Operators

Hyper-parameter values:

- Fraction of train concepts dropped in ConceptDrop: 0%, 5%, 10%, 20% (chosen: 5%)

Learning rate, batch size, regularization weights chosen from the original paper and executed using the implementation at: https://github.com/Tushar-N/attributes-as-operators.

---

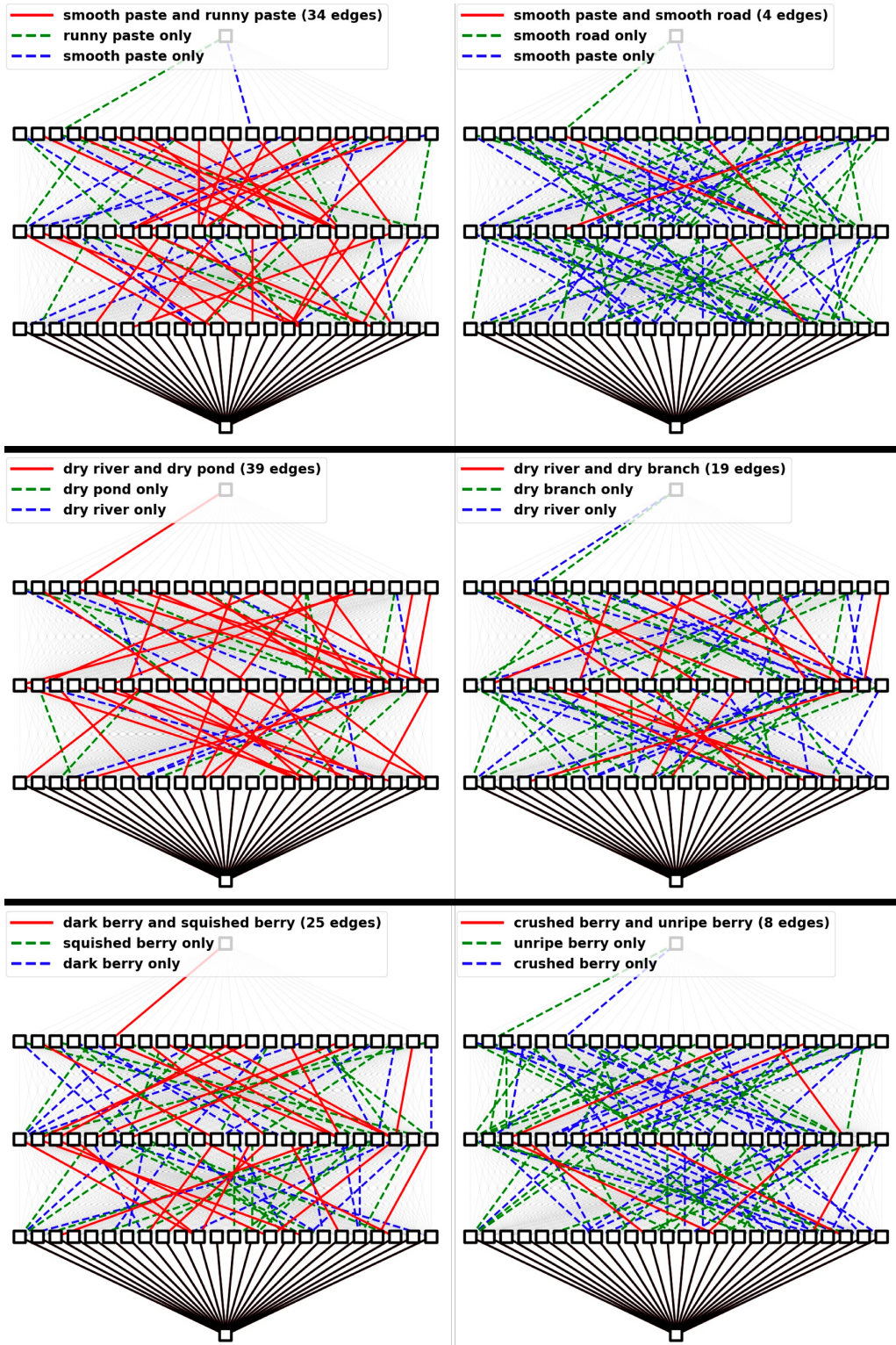# 2. Additional Topology Visualizations



Figure 1. Additional Examples of task driven topologies learned in TMN (similar to Figure 5 of the main text).