Neural Network Heuristic Functions: Taking Confidence into Account

Daniel Heller Patrick Ferber Julian Bitterwolf
Matthias Hein Jörg Hoffmann

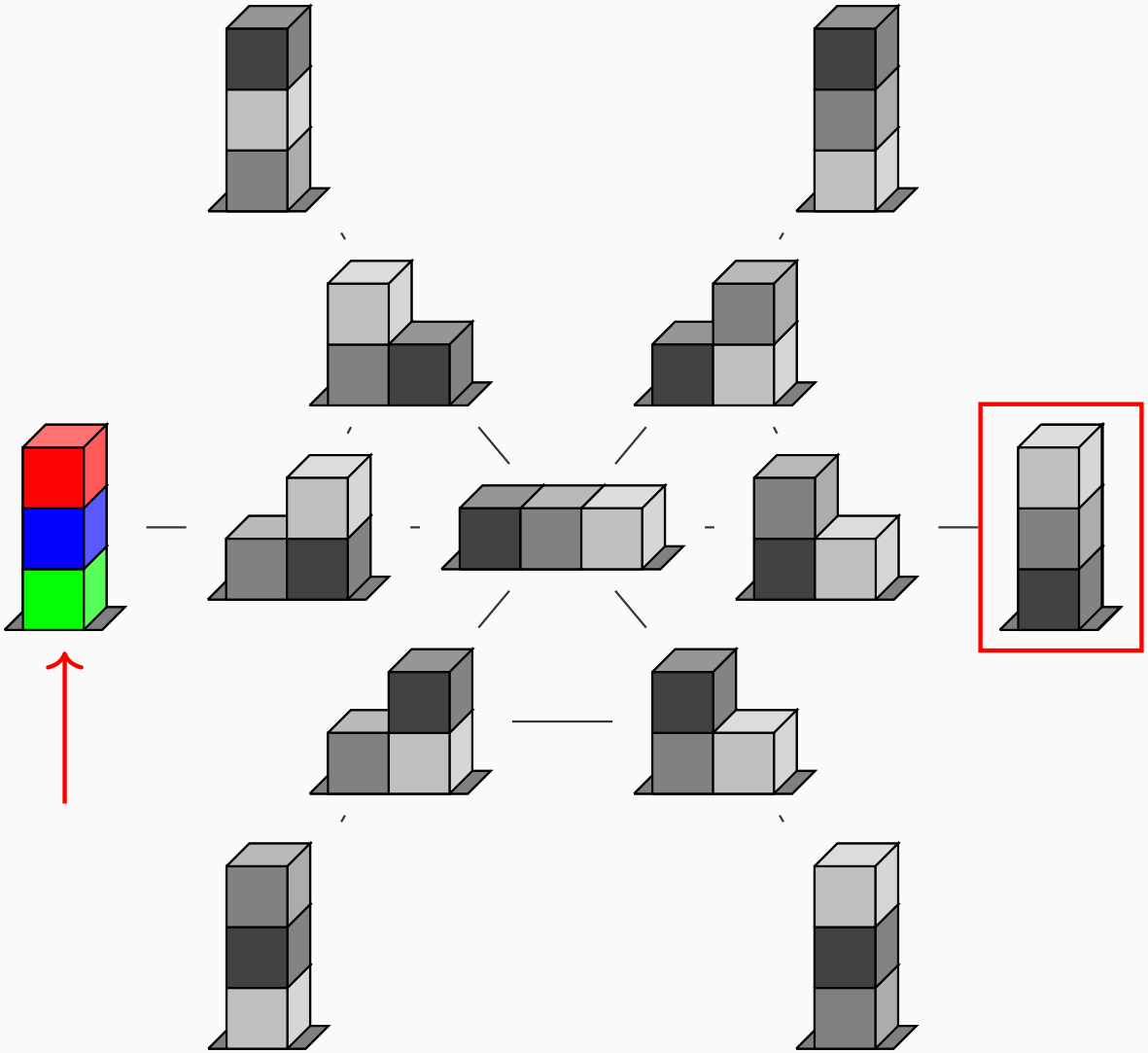




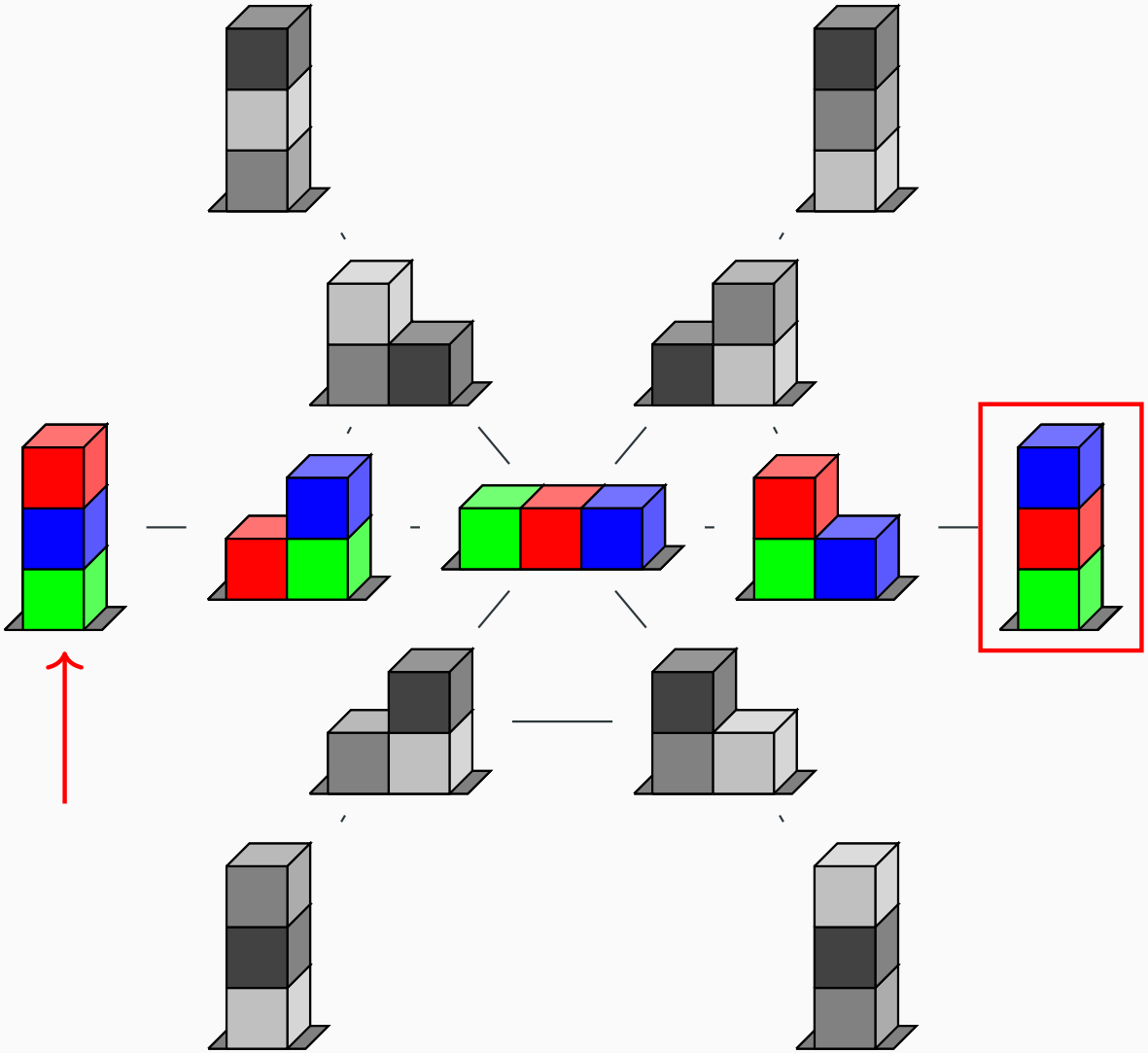
UNIVERSITÄT DES SAARLANDES
University of Basel
EBERHARD KARLS UNIVERSITÄT TÜBINGEN
German Research Center for Artificial Intelligence

15th International Symposium on Combinatorial Search

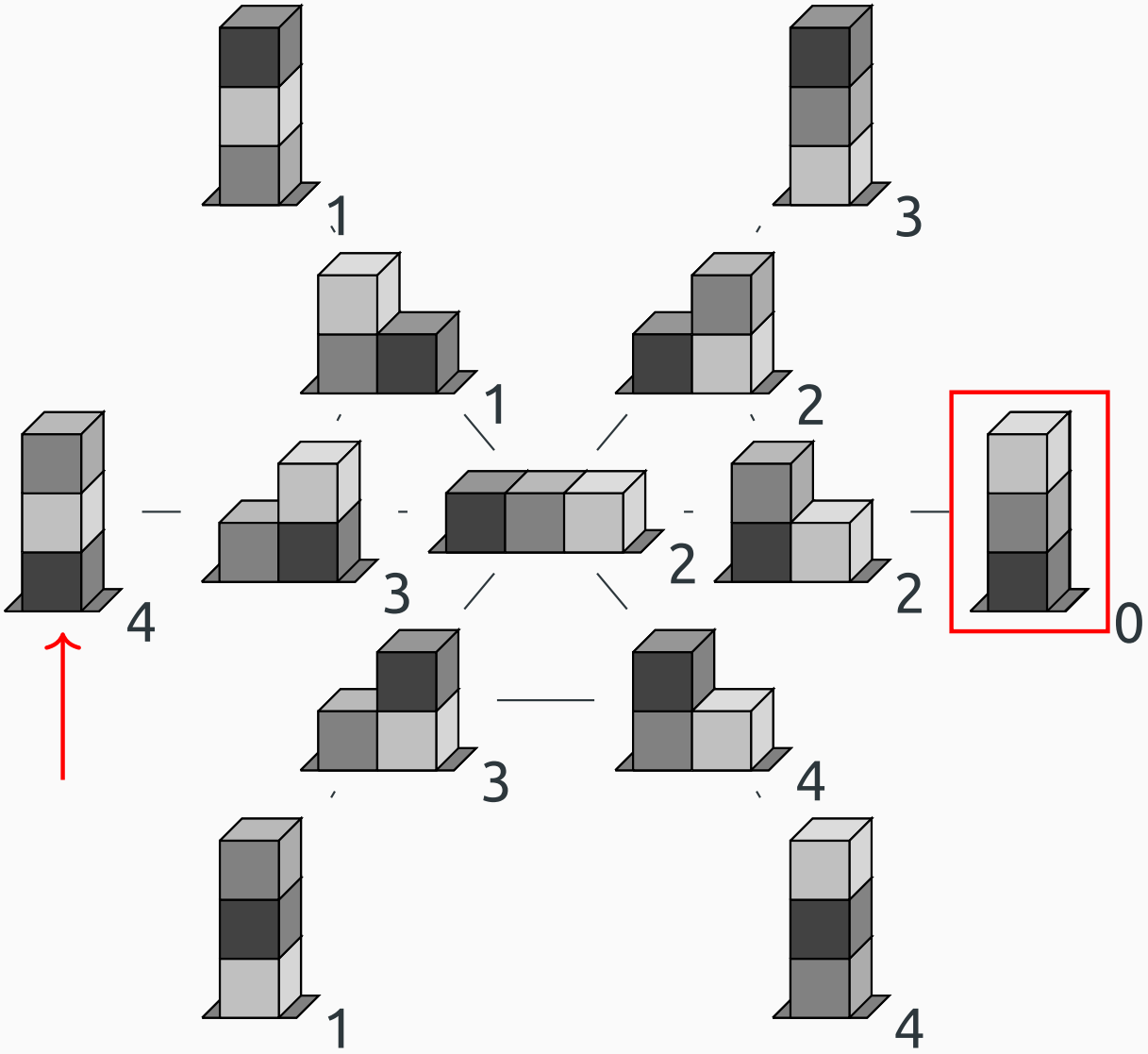
Heuristic Search



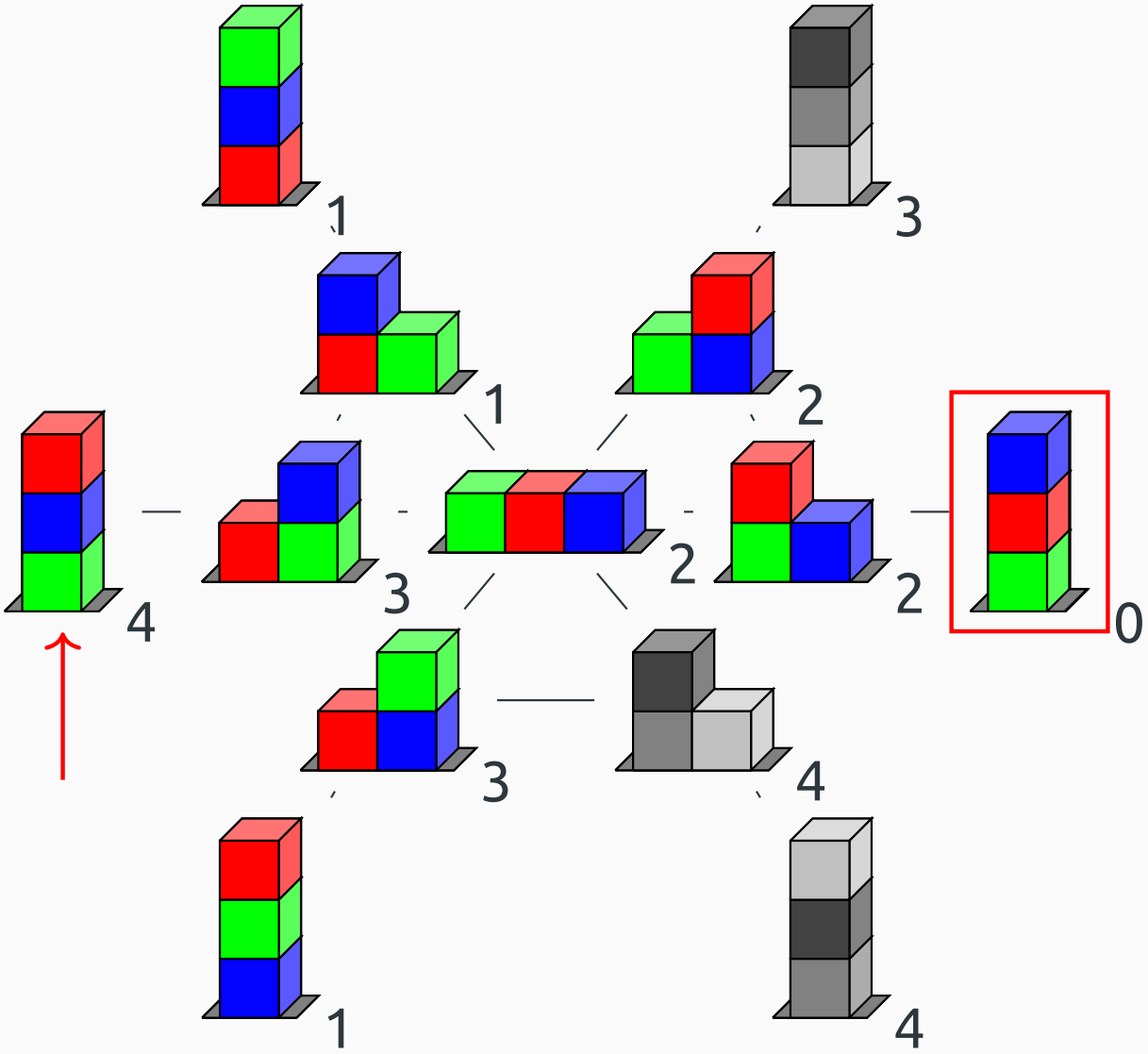
Heuristic Search



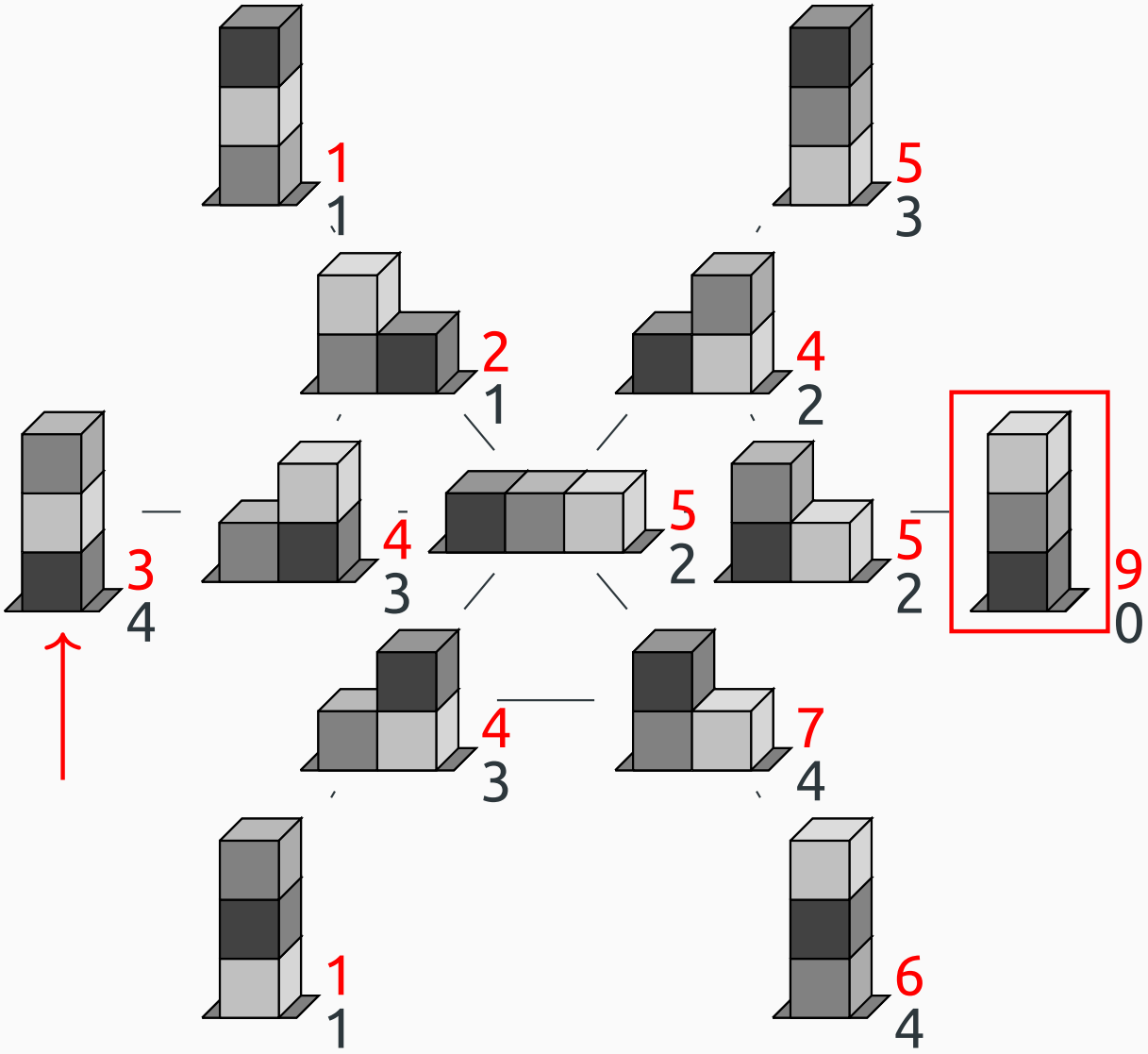
Heuristic Search



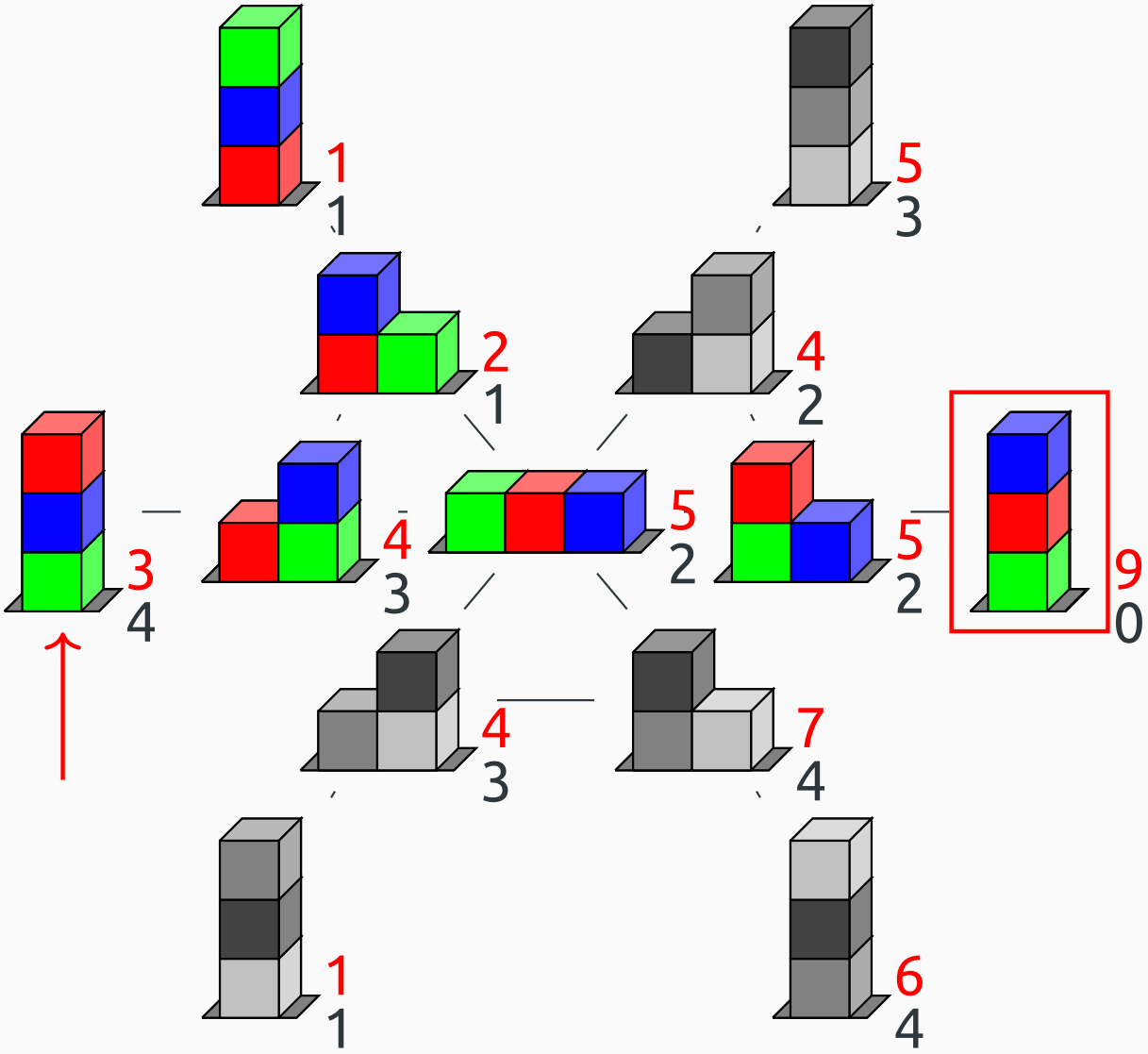
Heuristic Search



Heuristic Search



Heuristic Search



Contribution

Confidence-Aware Search Algorithms

- Mean Threshold
- Adaptive Threshold
- Prioritizing Queue

Out-Of-Distribution Training for Planning

- Uniform Noise
- Weighted Noise

Contribution

Confidence-Aware Search Algorithms

- Mean Threshold
- Adaptive Threshold
- Prioritizing Queue

Out-Of-Distribution Training for Planning

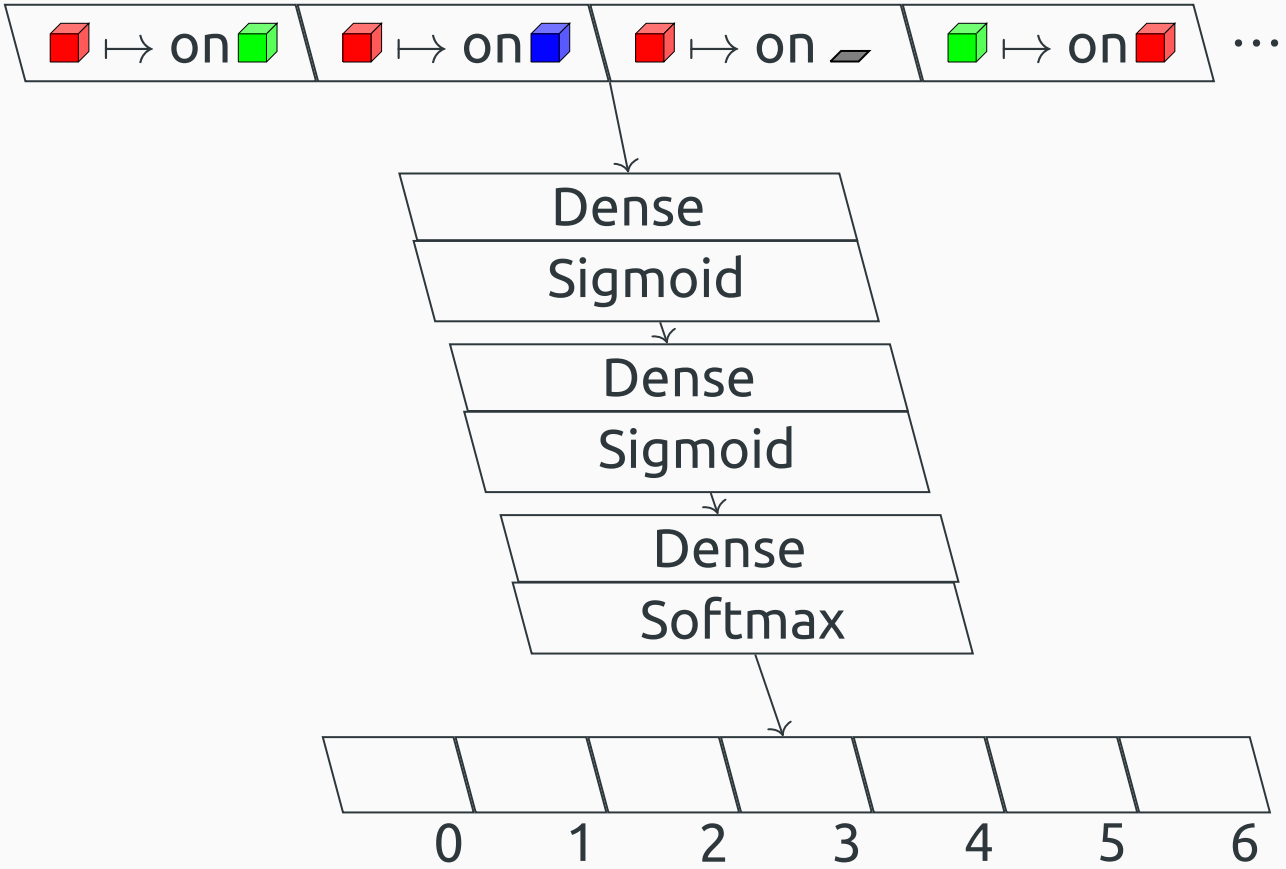
- Uniform Noise
- Weighted Noise

Heuristic

$$h : S \to \mathbb{R}_0^+ \cup \{\infty\}$$

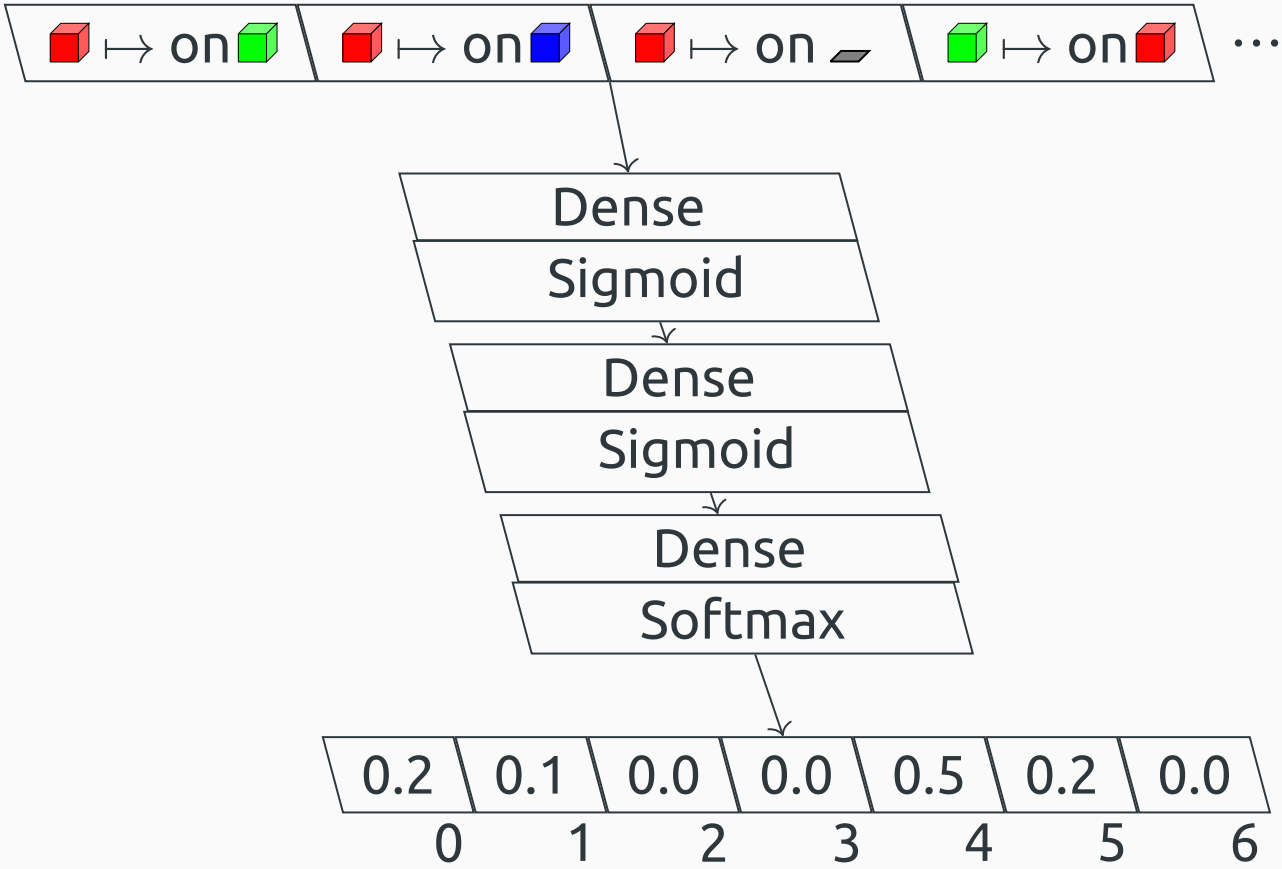
$$c : S \to [0, 1]$$

Neural Network Heuristic



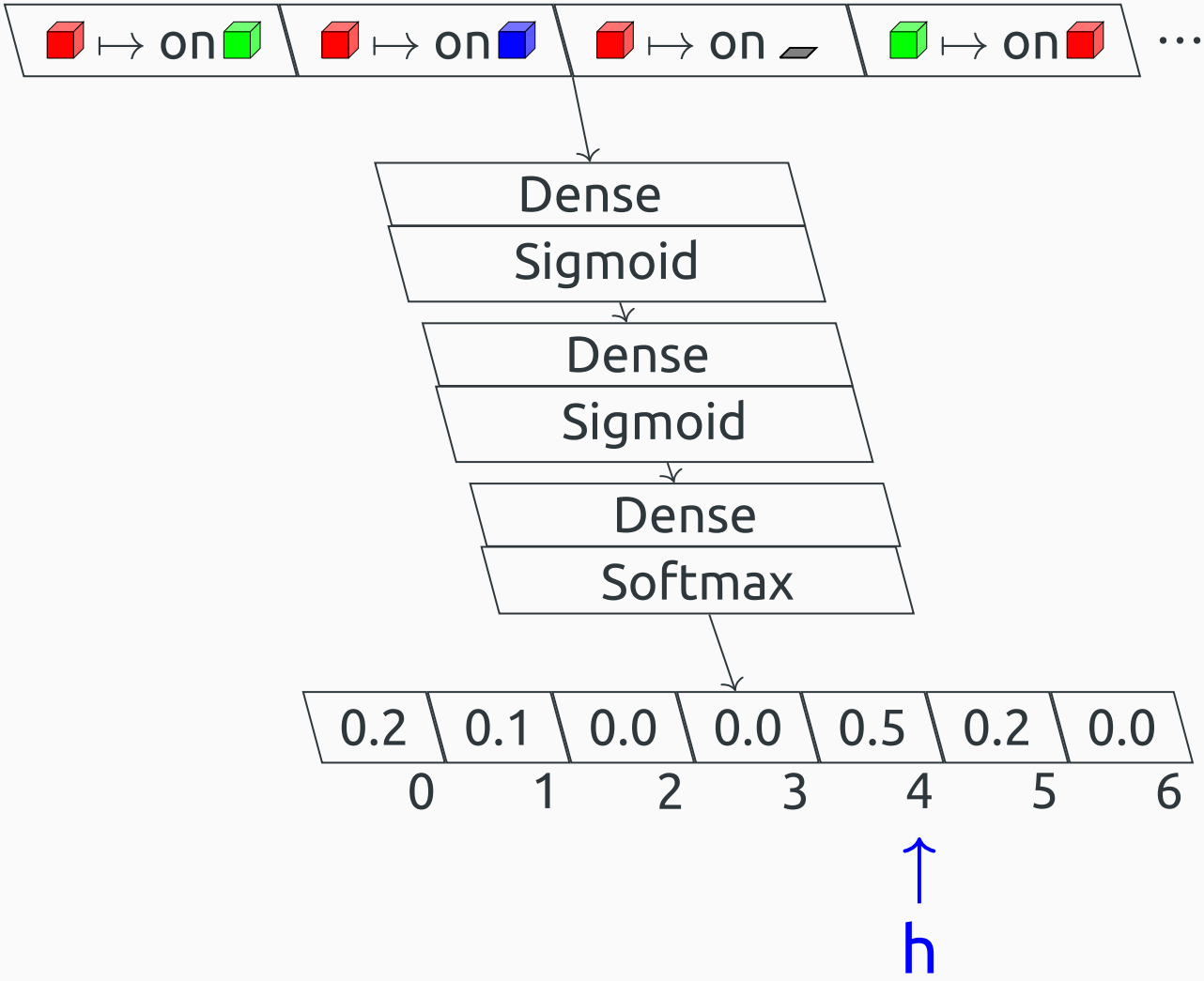
[Ferber *et al.*, 2020]

Neural Network Heuristic



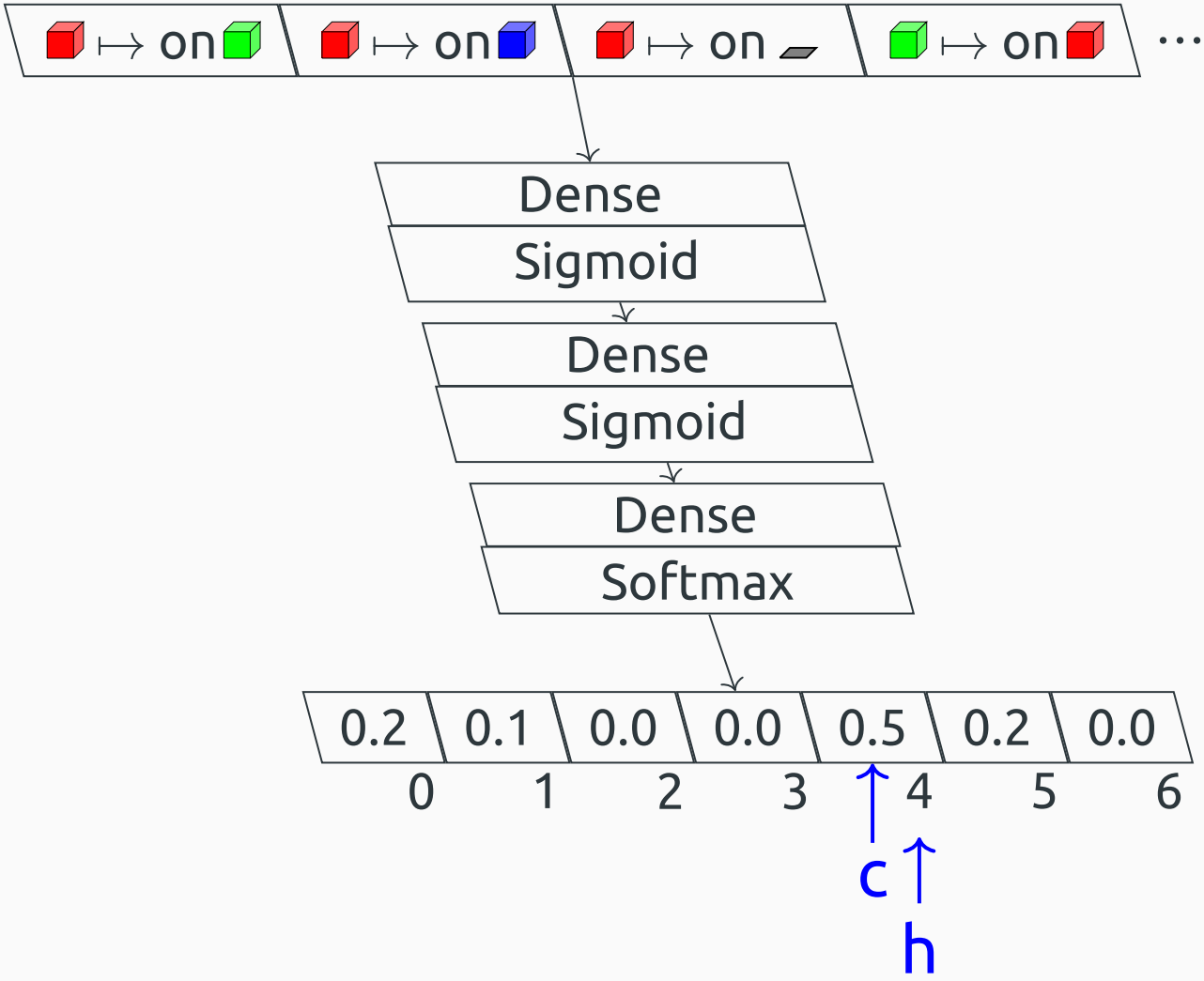
[Ferber *et al.*, 2020]

Neural Network Heuristic



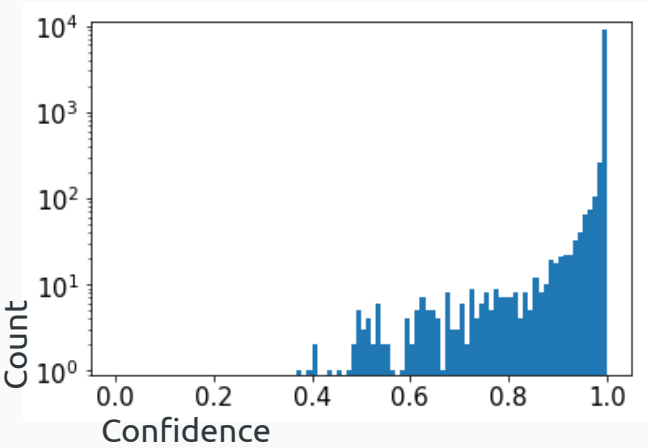
[Ferber *et al.*, 2020]

Neural Network Heuristic

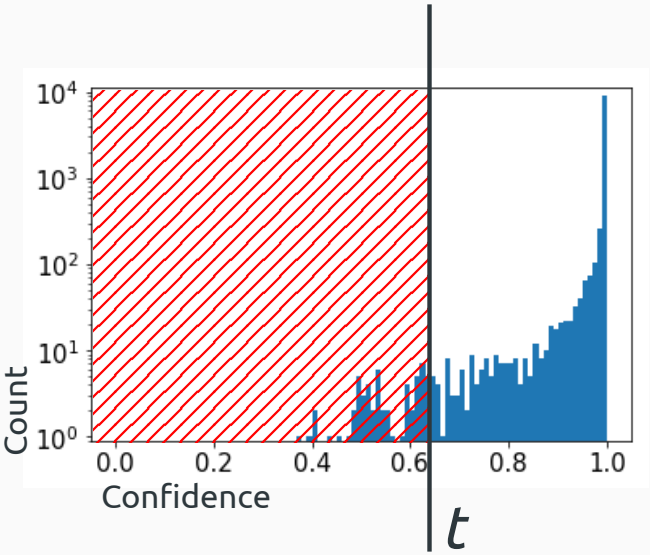


[Ferber *et al.*, 2020]

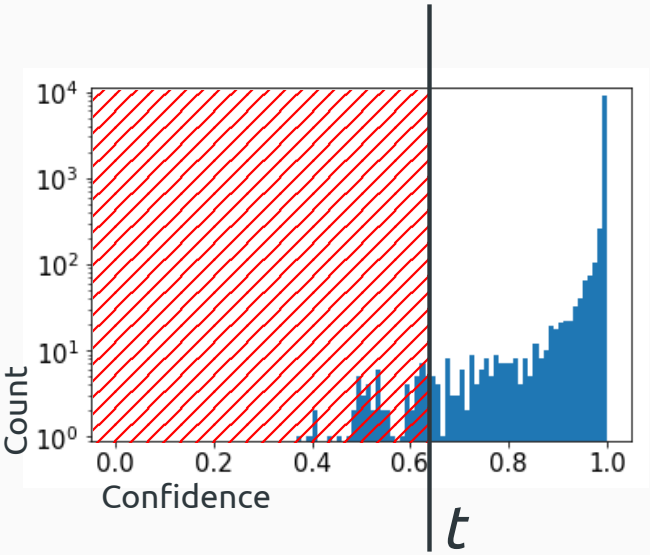
Mean Threshold



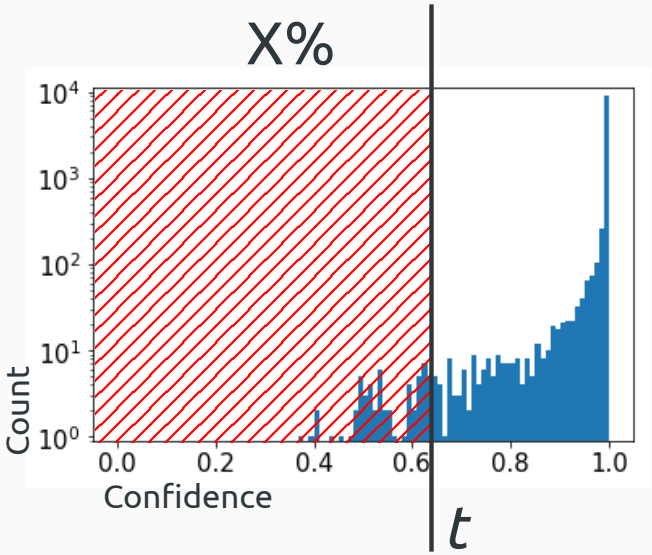
Mean Threshold



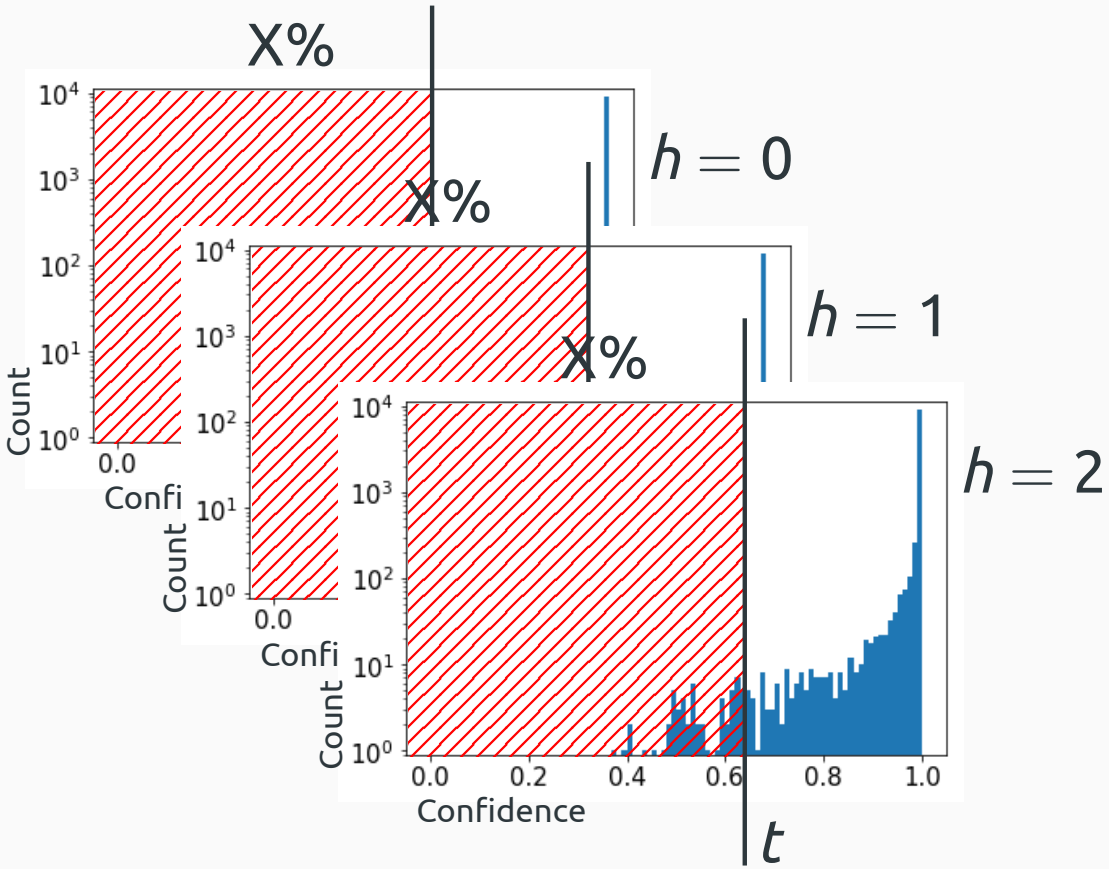
Mean Threshold



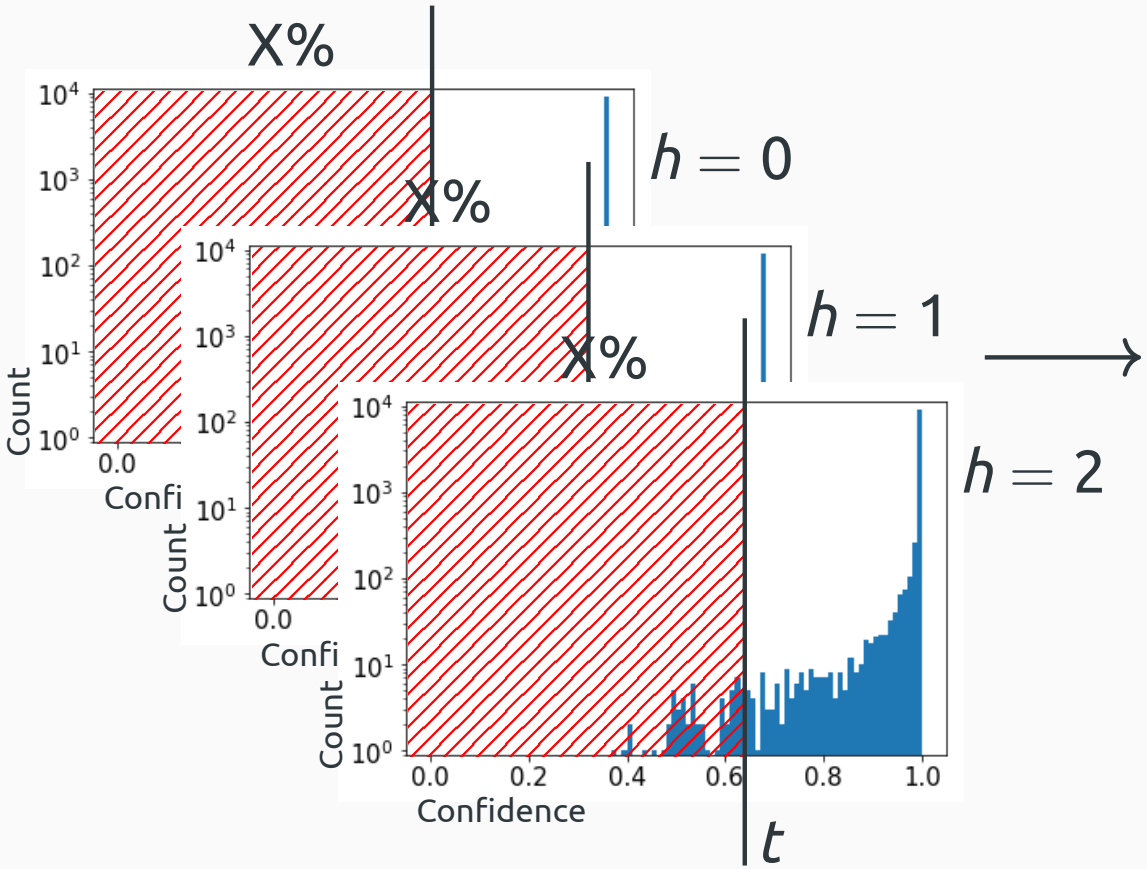
Mean Threshold

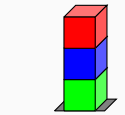


Adaptive Threshold



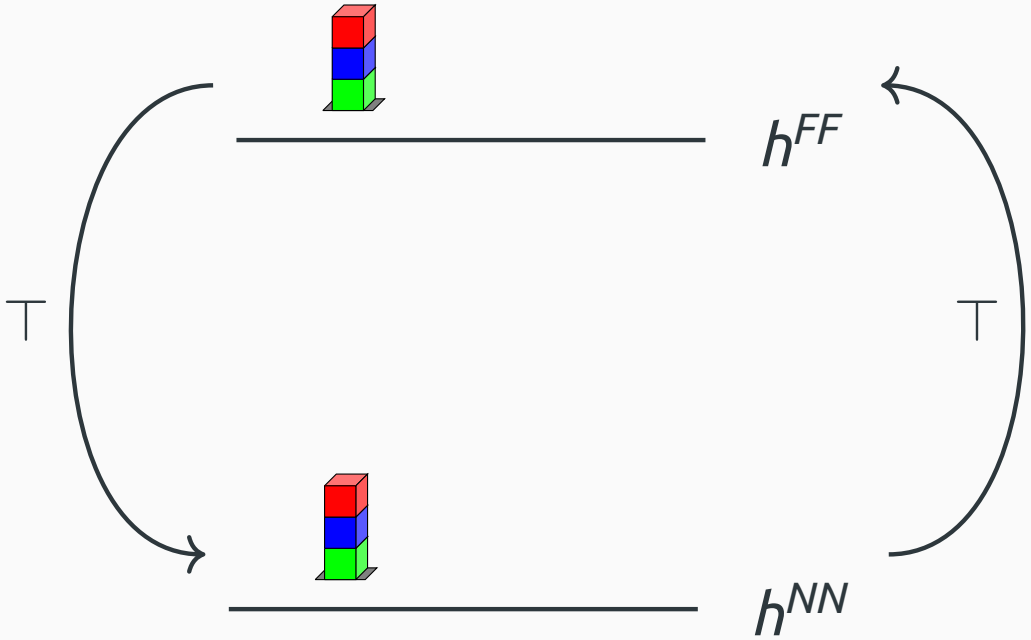
Adaptive Threshold



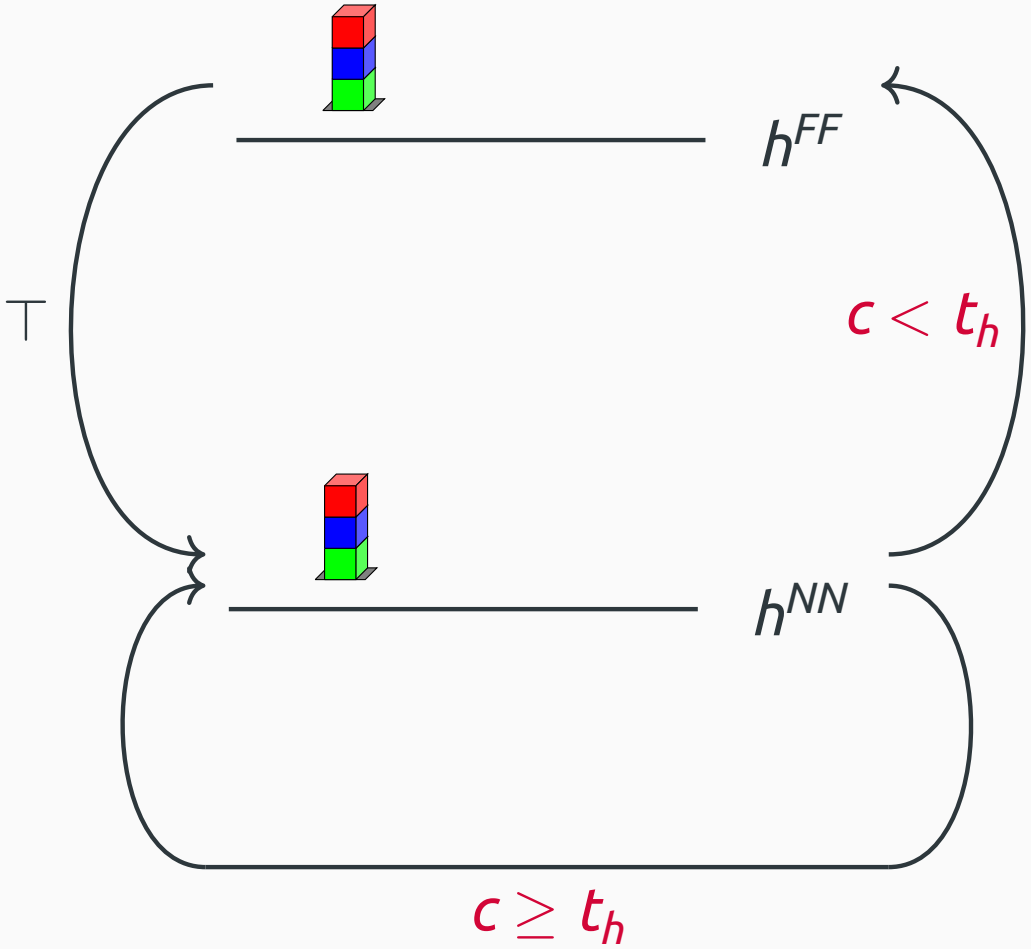


$h, c \geq t_h$

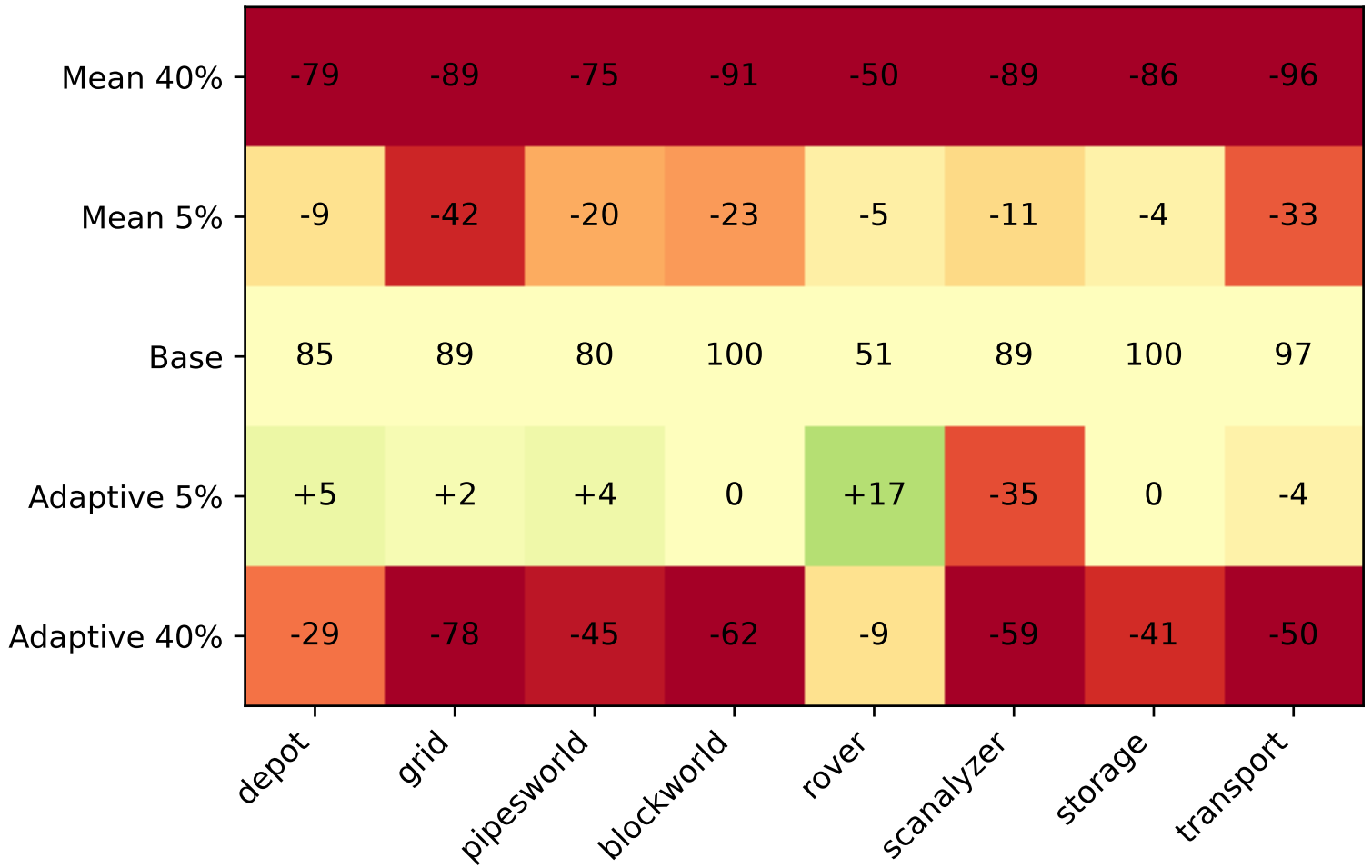
Prioritizing Queue



Prioritizing Queue



Effect on Coverage: Single Queue



Contribution

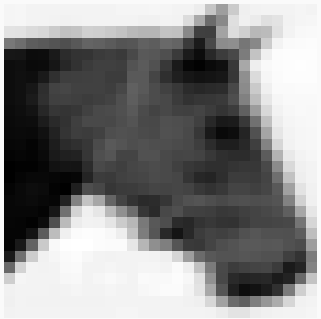
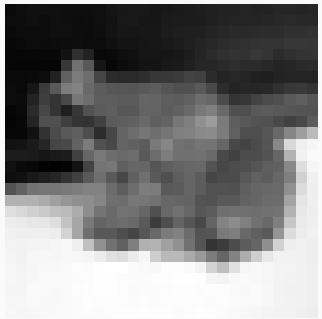
Confidence-Aware Search Algorithms

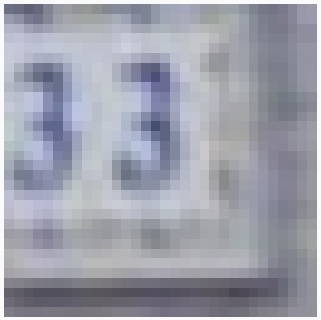
- Mean Threshold
- Adaptive Threshold
- Prioritizing Queue

Out-Of-Distribution Training for Planning

- Uniform Noise
- Weighted Noise

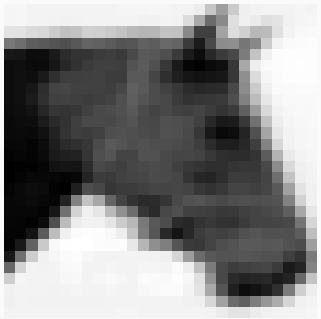
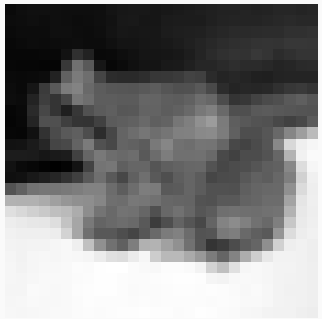
Overconfidence

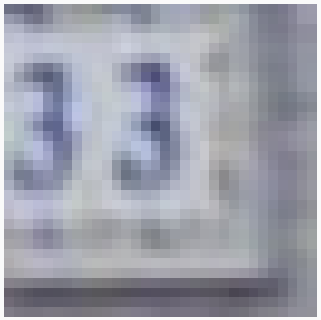


[Hein *et al.*, 2019]

Overconfidence

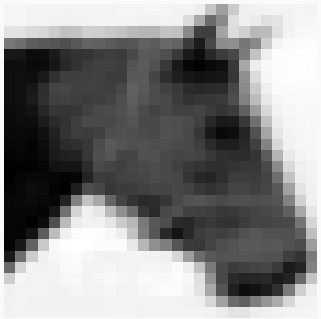




”2“ (99.6%)

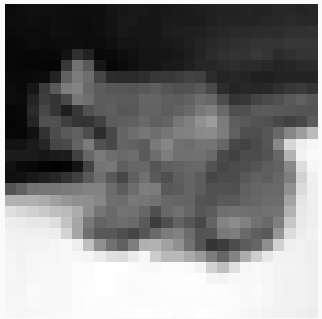


[Hein *et al.*, 2019]

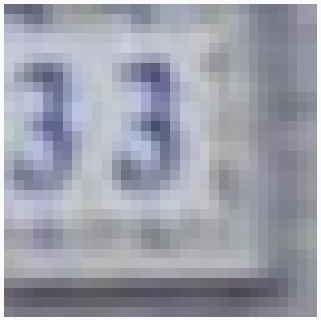
Overconfidence



”2“ (99.6%)

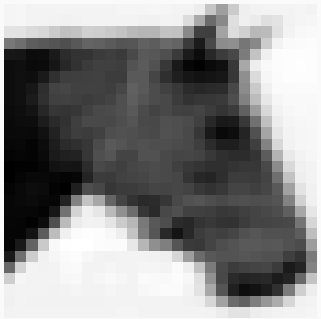


”2“ (99.1%)

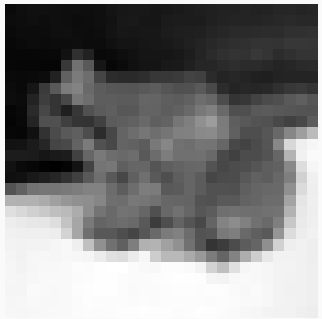


[Hein *et al.*, 2019]

Overconfidence

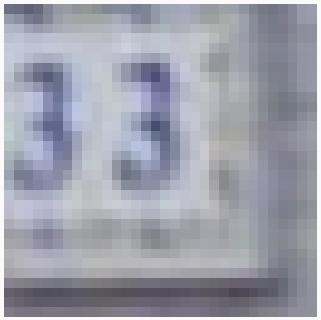


”2“ (99.6%)



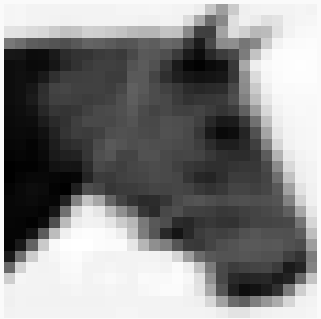
”2“ (99.1%)

”7“ (99.0%)

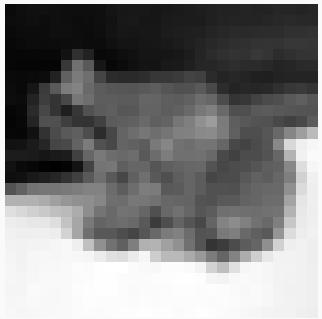


[Hein *et al.*, 2019]

Overconfidence

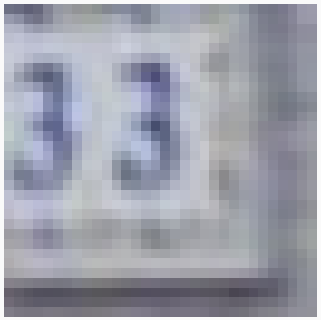


”2“ (99.6%)



”2“ (99.1%)

”7“ (99.0%)



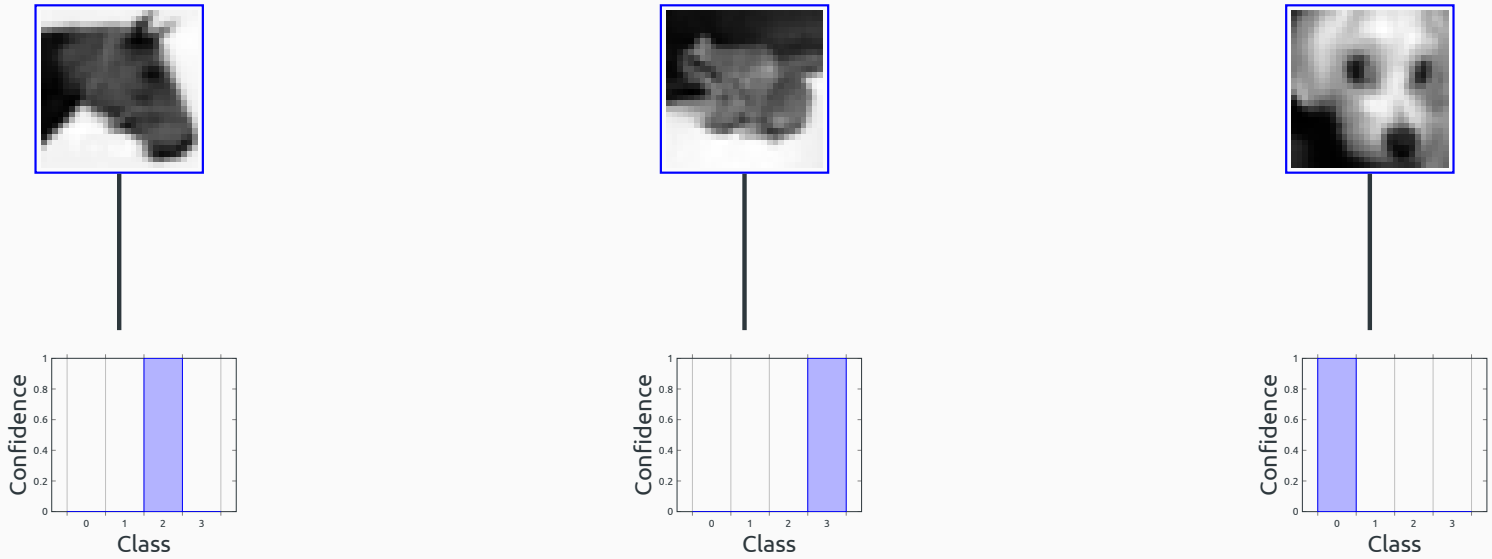
Dog (100.0%)

Bird (100.0%)

Airplane (100.0%)

[Hein *et al.*, 2019]

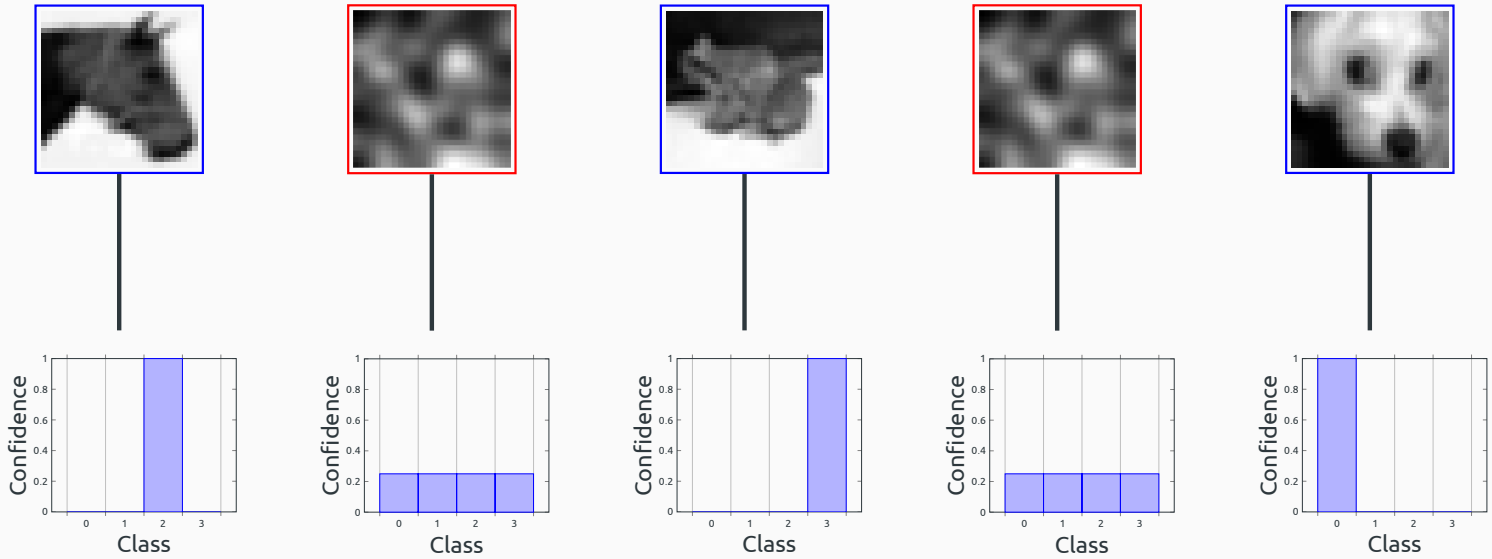
Out-Of-Distribution Training



$$\frac{1}{N}\sum_{i=1}^{N} L_{CE}(y_i, f(x_i))$$

[Hein *et al.*, 2019]

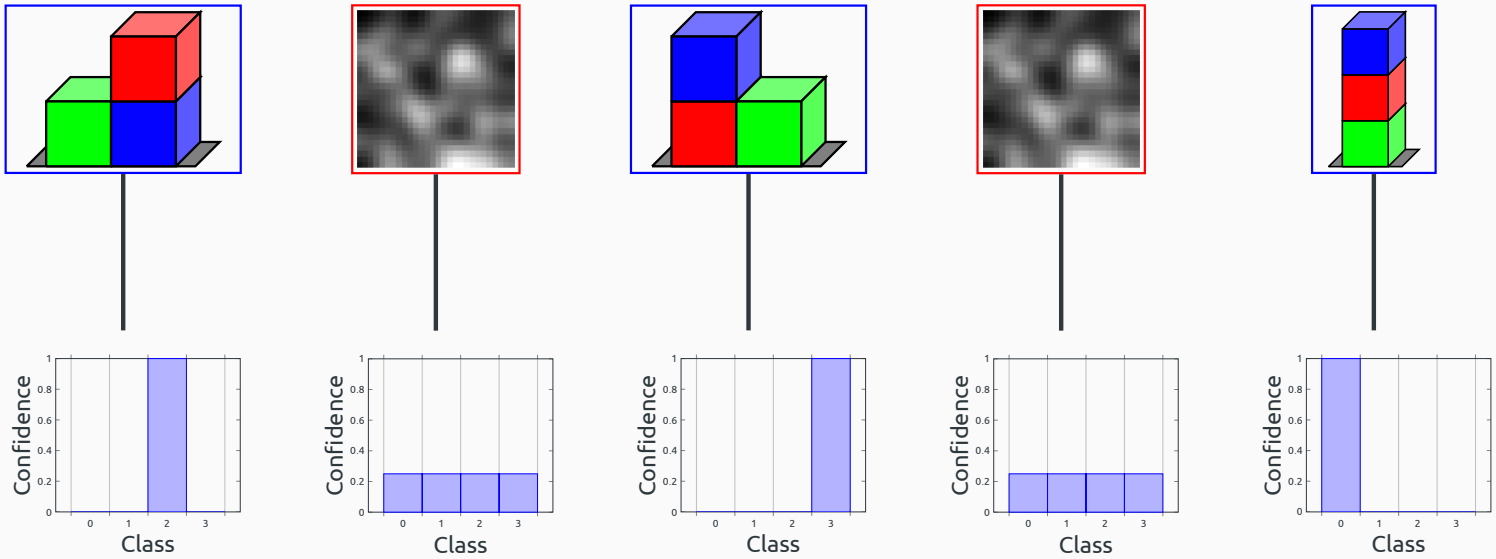
Out-Of-Distribution Training



$$\frac{1}{N}\sum_{i=1}^{N} L_{CE}(y_i, f(x_i)) + \lambda \max_k log(f(x_i)_k)$$

[Hein *et al.*, 2019]

Out-Of-Distribution Training

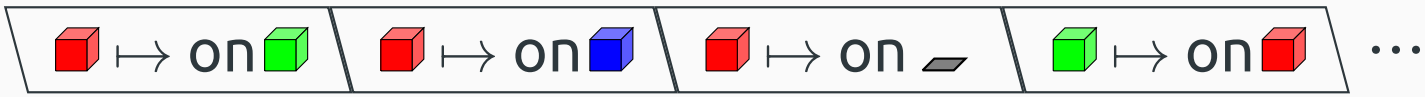


$$\frac{1}{N}\sum_{i=1}^{N} L_{CE}(y_i, f(x_i)) + \lambda \max_k log(f(x_i)_k)$$

[Hein *et al.*, 2019]

Out Distributions

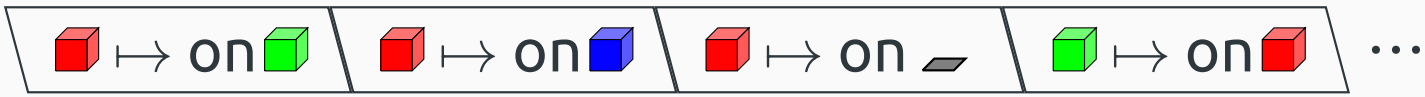
Uniform



Weighted

Out Distributions

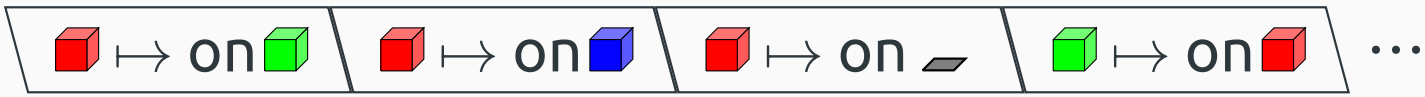
Uniform 50% 50% 50% 50%



Weighted

Out Distributions

Uniform 50% 50% 50% 50%



Weighted p_1 p_2 p_3 p_4

Effect on Confidence

Training Method	Confidence In	Confidence Uniform	Confidence Weighted	Ranking Coefficient
Standard	29.3	14.5	10.9	84.1
Uniform 50	30.0	0.5	12.0	84.0
Uniform 90	30.8	0.4	12.2	84.1
Weighted 50	29.0	10.9	1.1	84.6
Weighted 90	29.2	5.5	0.6	84.5

Effect on Confidence

Training Method	Confidence In	Confidence Uniform	Confidence Weighted	Ranking Coefficient
Standard	29.3	14.5	10.9	84.1
Uniform 50	30.0	0.5	12.0	84.0
Uniform 90	30.8	0.4	12.2	84.1
Weighted 50	29.0	10.9	1.1	84.6
Weighted 90	29.2	5.5	0.6	84.5

Effect on Confidence

Training Method	Confidence In	Confidence Uniform	Confidence Weighted	Ranking Coefficient
Standard	29.3	14.5	10.9	84.1
Uniform 50	30.0	0.5	12.0	84.0
Uniform 90	30.8	0.4	12.2	84.1
Weighted 50	29.0	10.9	1.1	84.6
Weighted 90	29.2	5.5	0.6	84.5

Effect on Confidence

Training Method	In	Confidence Uniform	Confidence Weighted	Ranking Coefficient
Standard	29.3	14.5	10.9	84.1
Uniform 50	30.0	0.5	12.0	84.0
Uniform 90	30.8	0.4	12.2	84.1
Weighted 50	29.0	10.9	1.1	84.6
Weighted 90	29.2	5.5	0.6	84.5

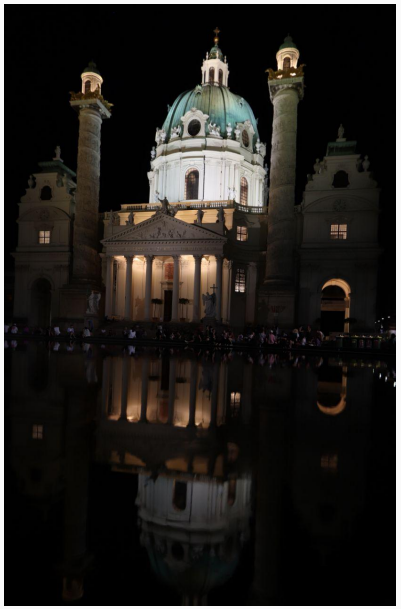
Effect on Confidence

Training Method	In	Confidence Uniform	Confidence Weighted	Ranking Coefficient
Standard	29.3	14.5	10.9	84.1
Uniform 50	30.0	0.5	12.0	84.0
Uniform 90	30.8	0.4	12.2	84.1
Weighted 50	29.0	10.9	1.1	84.6
Weighted 90	29.2	5.5	0.6	84.5

Effect on Coverage: Noise

Summary

- Confidence-awareness improves search
- Adapt out-of-distribution training for planning
- Improving confidence did not yet improve search



References

Patrick Ferber, Malte Helmert, and Jörg Hoffmann. Neural network heuristics for classical planning: A study of hyperparameter space. In Giuseppe De Giacomo, editor, *Proceedings of the 24th European Conference on Artificial Intelligence (ECAI 2020)*, pages 2346–2353. IOS Press, 2020.

Matthias Hein, Maksym Andriushchenko, and Julian Bitterwolf. Why relu networks yield high-confidence predictions far away from the training data and how to mitigate the problem. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 41–50, 2019.