Deep Learning AIE:

Implementing W&B Sweeps into ConvNets Lesson

Simeon Betapudi, Jayden Cruz, Evan Dant

What is sweeps??

Def'n

Sweeps is an automated experimentation framework in W&B that systematically test different hyperparameter combinations

Analogy

Instead of manually guessing hyperparameters.
It tries many configs with minimal manual effort

Benefits

- Efficient
- Reproducible
- Gives better performance insights
- Helps find ideal model settings faster

Our Goal/Idea

- Find the fastest & most computationally efficient way to search for best configurations
- Goal: Achieve at least 75% accuracy
- Reduce train time without totally sacrificing accuracy

Motivation

- Playing around with hyperparameters by hand is slow and random

- We know that training for more epochs usually improves accuracy, but we limited ourselves to just 5 epochs → needed the best parameters early
- To decrease train time, increased batch size from 128 to 1024

Dataset (CIFAR-10)

- 60,000 images (32x32, RGB)
- 10 classes (airplane, car, cat, dogs,etc.)
- Standard benchmark for image classification

Methodology (How We Did It)

- **Model:** Convolutional Neural Network (CNN)
- **Sweeps:** W&B grid search
- **Config parameters:** base_channels, channel_mult, n_conv_layers, kernel_size, stride
- **Activation: Leaky ReLU**

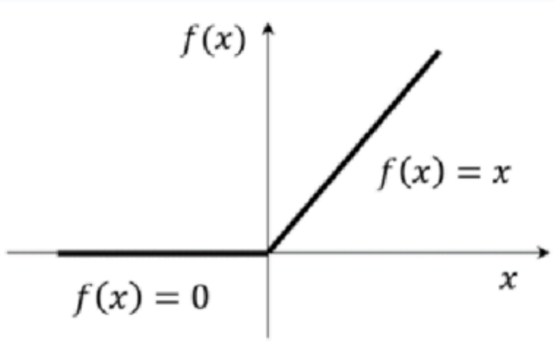
Our Sweep Config File

- Base Channels → starting number of filters in conv layers
- Channel Multiplier → scales number of filters in deeper layers
- Number of Convolution Layers → network depth
- Kernel Size → filter size for feature extraction
- Stride → step size of filters

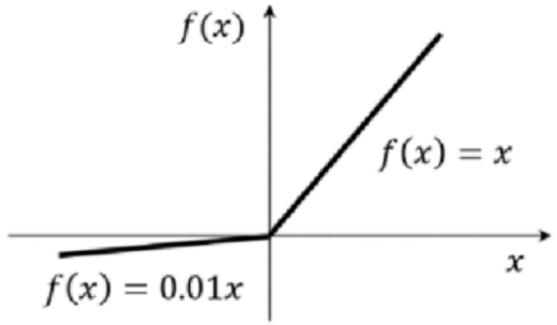
base_channels: [32, 64]
channel_mult: [2, 3]
n_conv_layers: [4, 5]
kernel_size: [3, 5]
stride: [1, 2, 3]

48 total combinations

Why Leaky ReLU?



ReLU activation function



LeakyReLU activation function

01

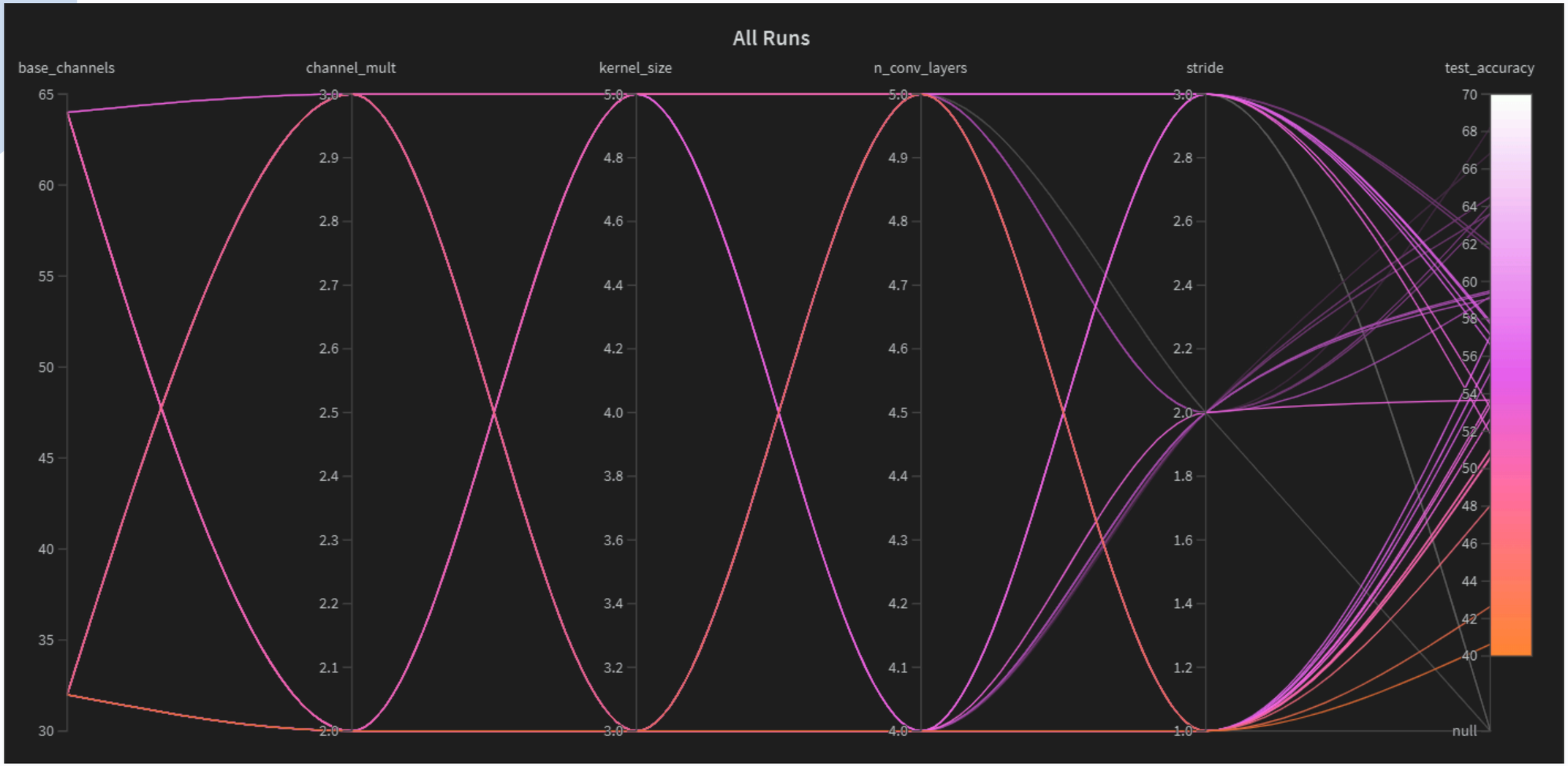
Standard ReLu can cause “dead neurons” output stuck at 0, never recover

02

Leaky ReLu gives a small slope on the negative inputs → prevents neurons from dying, improves gradient flow

Results

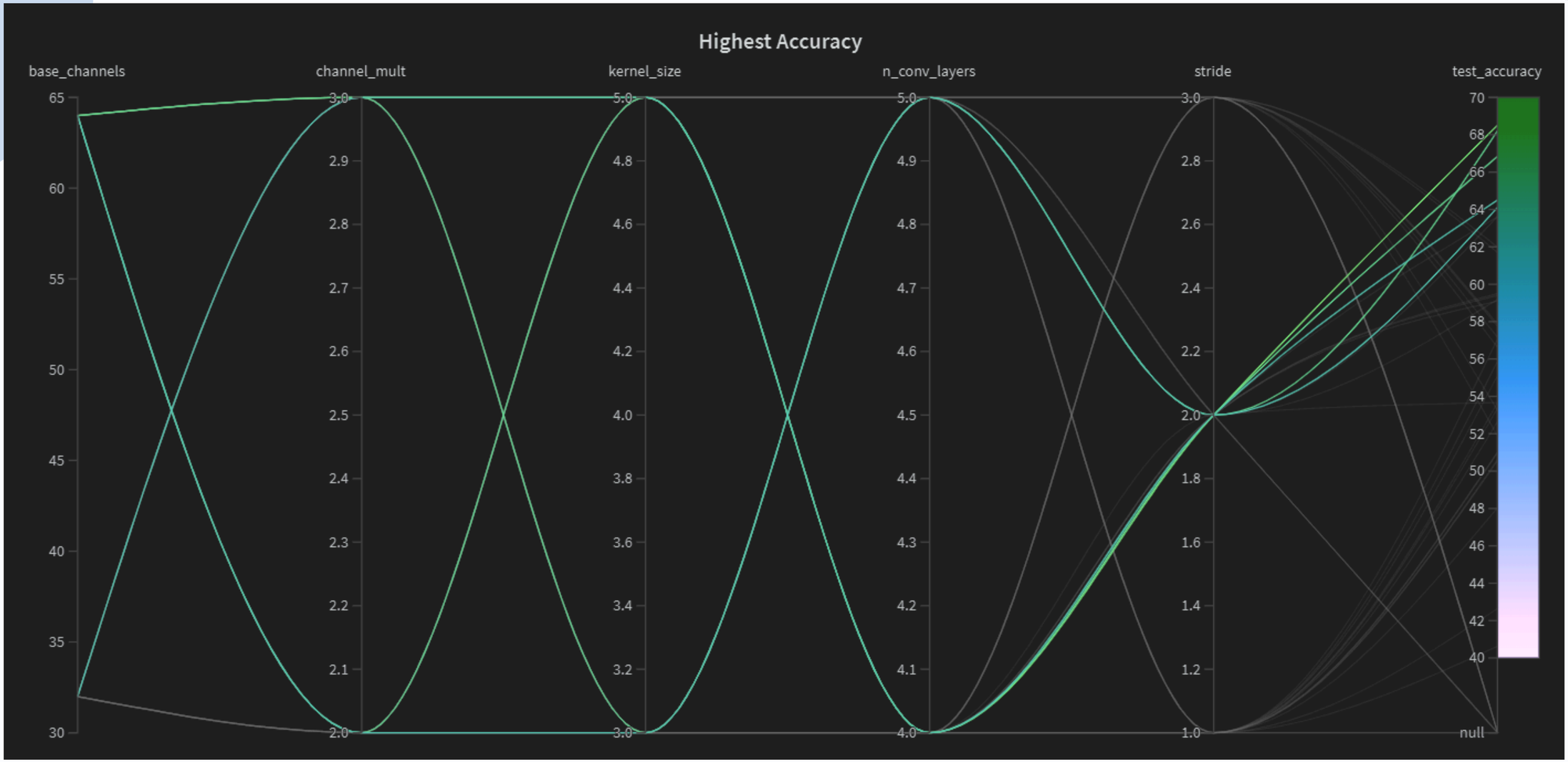
All Runs


All Runs
base_channels
channel_mult
kernel_size
n_conv_layers
stride
test_accuracy
65
60
55
50
45
40
35
30
3.0
2.9
2.8
2.7
2.6
2.5
2.4
2.3
2.2
2.1
2.0
5.0
4.8
4.6
4.4
4.2
4.0
3.8
3.6
3.4
3.2
3.0
5.0
4.9
4.8
4.7
4.6
4.5
4.4
4.3
4.2
4.1
4.0
3.0
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
70
68
66
64
62
60
58
56
54
52
50
48
46
44
42
40
null

Highest Accuracy Runs

	Base Channels	Channel Multiplier	Layers	Kernel Size	Stride	Accuracy	Run Times (s)
1.	64	3	4	5	2	68.52%	71.9
2.	64	3	5	3	2	68.22%	74.3
3.	64	2	4	5	2	66.85%	69.1
4.	32	3	4	5	2	64.51%	69.2
5.	64	2	5	3	2	64.11%	69.9

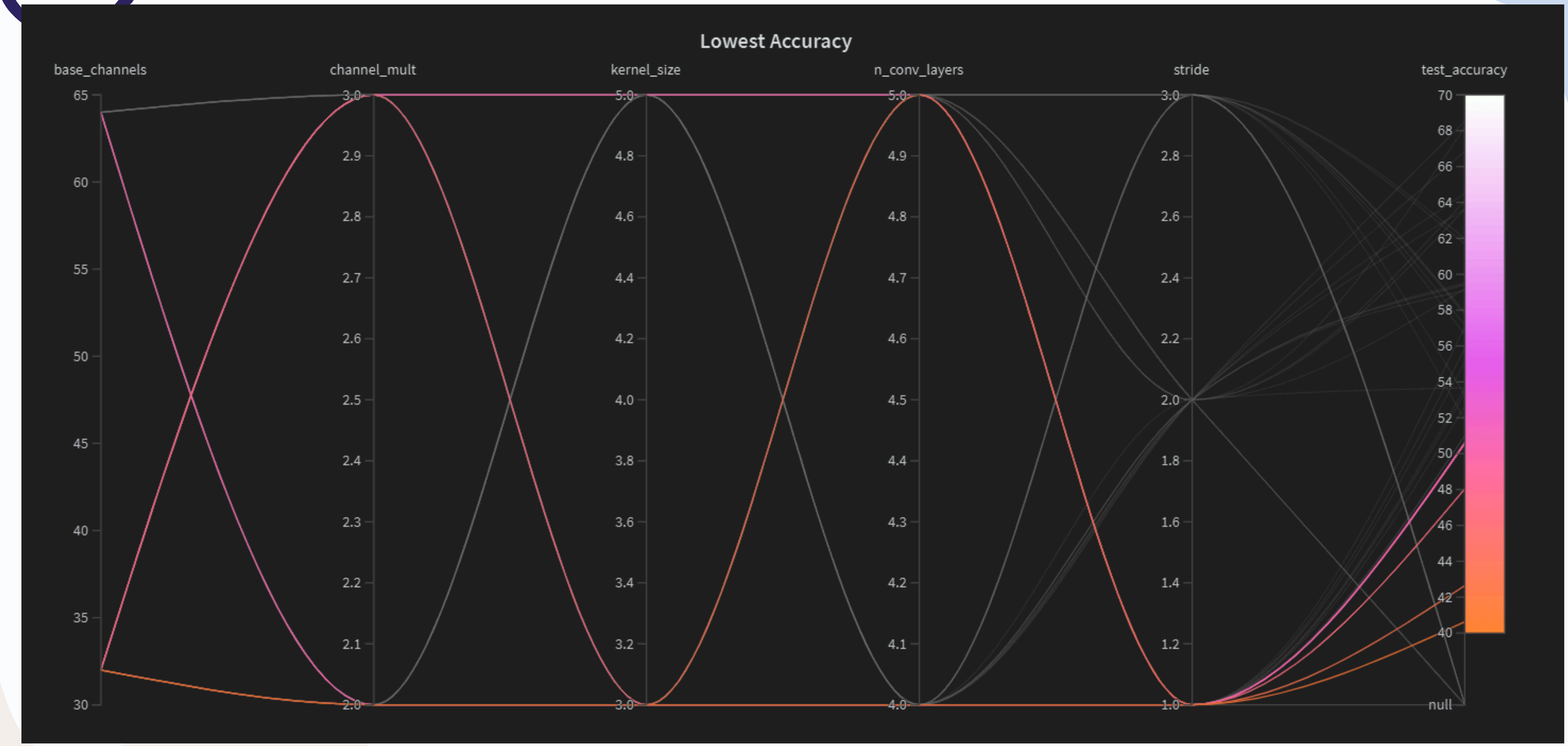
Highest Accuracy Runs



Lowest Accuracy Runs

	Base Channels	Channel Multiplier	Layers	Kernel Size	Stride	Accuracy	Run Time (s)
29.	64	2	4	3	1	50.63%	100.9
30.	32	3	5	5	1	50.56%	584.1
31.	32	3	4	3	1	48.04%	123.7
32.	32	2	5	3	1	42.65%	101.0
33.	32	2	4	3	1	40.63%	81.2

Lowest Accuracy Runs



Longest Run Times

	Base Channels	Channel Multiplier	Layers	Kernel Size	Stride	Run Time (s)	Accuracy
1.	32	3	5	5	1	584	50.56%
2.	32	3	5	3	1	459	51.00%
3.	64	3	4	3	1	245	48.04%
4.	64	2	5	5	1	201	42.65%
5.	64	2	5	3	1	173	40.63%

Shortest Run Times

	Base Channels	Channel Multiplier	Layers	Kernel Size	Stride	Run Time (s)	Accuracy
1.	32	2	4	3	2	67.6	53.67%
2.	32	2	5	3	3	67.7	51.87%
3.	32	3	4	3	3	67.9	57.15%
4.	32	2	4	3	3	68.0	53.25%
5.	32	2	4	5	2	68.1	59.39%

Run Times

Longest: Base Channels: 32, Channel Multiplier: 3, Layers: 5, Kernel Size: 5, Stride: 1
584secs = ~10mins

Shortest: Base Channels: 32, Channel Multiplier: 2, Layers 4, Kernel Size 3, Stride 2
67secs = ~1.07mins

Conclusion

01 Sweep let us automate hyperparameter tuning while letting us explore the hyperparameter space

02 Longer runs != more accuracy

03 Stride has a large impact on efficiency and accuracy

04 Future Improvements: Larger models, more hyperparameters

References & Resources

- **Dr. Hawley's Course Lesson - CNNs:** https://github.com/drscotthawley/DLAIE/blob/main/Lessons/06_ConvNets.ipynb

- **W&B Sweeps Documentation:** https://docs.wandb.ai/guides/sweeps/
- **More info about our Dataset**: https://www.cs.toronto.edu/%7Ekriz/cifar.html