Image Classification with CIFAR-10

April 2023

Presenters: Snithika Reddy, Hsi-sheng Wei, Andrew Abrahamian

Meet Our Team

Snithika Kalakoti
Data Scientist

Hsi-sheng Wei
Data Scientist

Andrew Abrahamian
Data Scientist

Agenda

Framing the Objective: The What, How, and Why?

Modeling Approach & Experiments

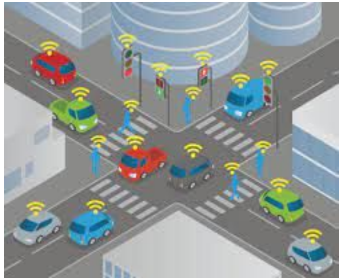
Results & Conclusions

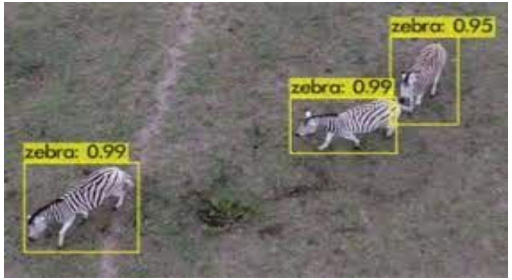
Ethical Considerations & Limitations

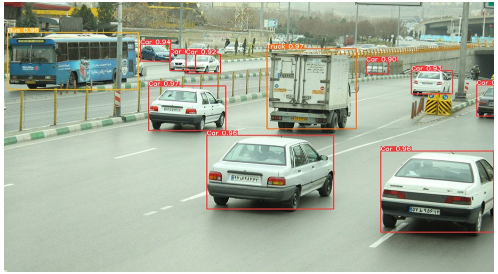
Q&A

What, How, and Why?:

Startup that builds machine learning models to classify images

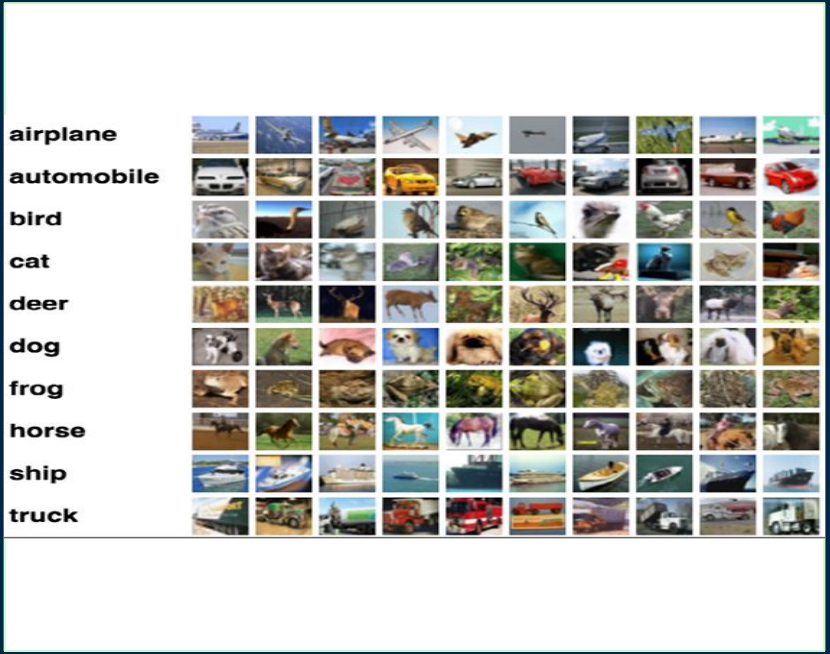






Our Data & Solution Approach

CIFAR-10 Dataset



FNN Confusion Matrix

True Label \ Predicted Label	airplane	automobile	bird	cat	deer	dog	frog	horse	ship	truck
airplane	452	43	117	19	37	5	50	19	192	66
automobile	16	634	30	16	12	2	27	18	85	160
bird	63	24	427	80	117	35	154	43	37	20
cat	15	23	138	336	73	66	213	41	36	59
deer	28	14	208	49	396	13	172	59	39	22
dog	6	22	179	233	67	233	136	54	41	29
frog	2	9	97	55	68	21	696	18	15	19
horse	17	17	131	84	106	25	56	491	21	52
ship	72	65	23	24	21	4	23	15	687	66
truck	23	184	30	32	8	7	41	36	60	579

- CIFAR-10: 60k 32x32 images (50k training, 10k testing): ideal for our use case
- Starter FNN model did not perform well (49% test accuracy); animal precision was a problem

Building Deeper CNN Models

CNN1 (70% test accuracy)

True Label \ Predicted Label	airplane	automobile	bird	cat	deer	dog	frog	horse	ship	truck
airplane	765	32	66	18	10	3	8	11	56	31
automobile	17	873	3	9	1	0	5	2	16	74
bird	61	14	679	55	36	53	50	27	13	12
cat	22	27	106	492	56	135	63	48	18	33
deer	30	8	126	76	570	30	61	77	11	11
dog	15	14	100	195	34	525	25	67	9	16
frog	8	19	64	55	24	22	763	14	13	18
horse	18	4	60	31	43	42	11	766	2	23
ship	71	38	16	10	2	4	1	4	826	28
truck	33	116	11	11	1	5	3	11	31	778

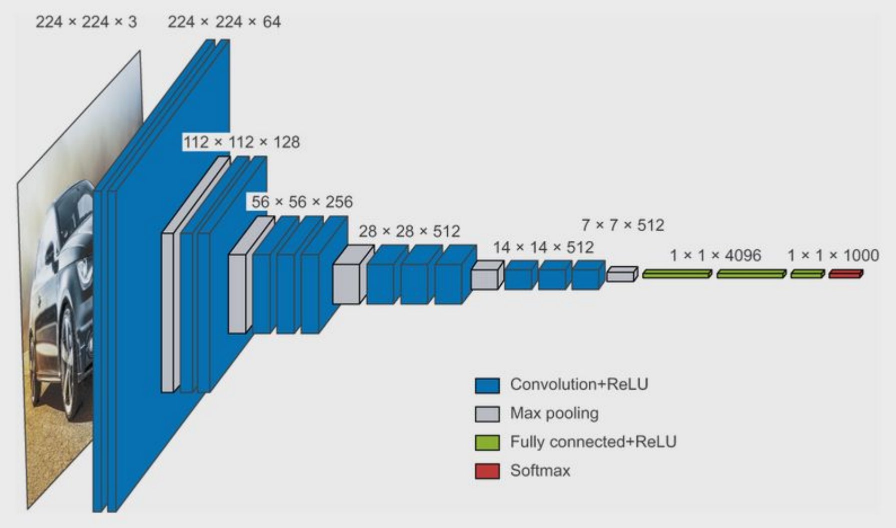
CNN3 (83% test accuracy)

True Label \ Predicted Label	airplane	automobile	bird	cat	deer	dog	frog	horse	ship	truck
airplane	783	5	88	27	16	8	9	7	46	11
automobile	5	918	1	4	2	4	7	1	19	39
bird	14	0	796	58	47	45	27	5	6	2
cat	7	1	40	731	30	124	48	11	6	2
deer	2	1	49	46	833	30	24	12	3	0
dog	4	1	23	154	22	769	11	11	4	1
frog	2	1	29	43	11	8	898	2	5	1
horse	10	1	25	43	30	53	6	824	2	6
ship	24	3	10	10	2	1	4	1	935	10
truck	22	43	6	14	3	4	6	3	20	879

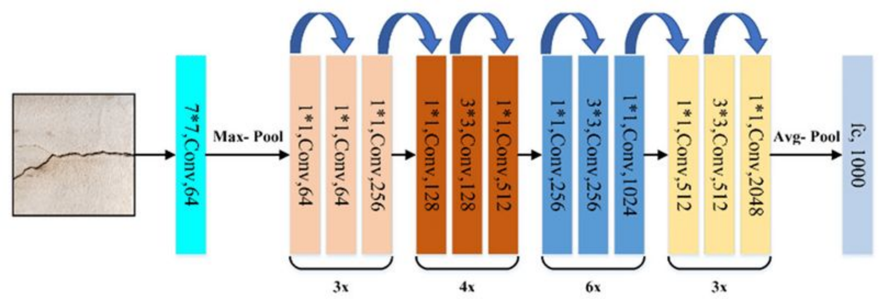
- Additional Batch Normalization, Dense, & Dropout layers
- Better precision with non-animal classes
- Challenge to differentiate between cats and dogs

Transfer Learning with VGG16 and ResNet50 Pre-Trained Models

VGG16 (86% test accuracy)

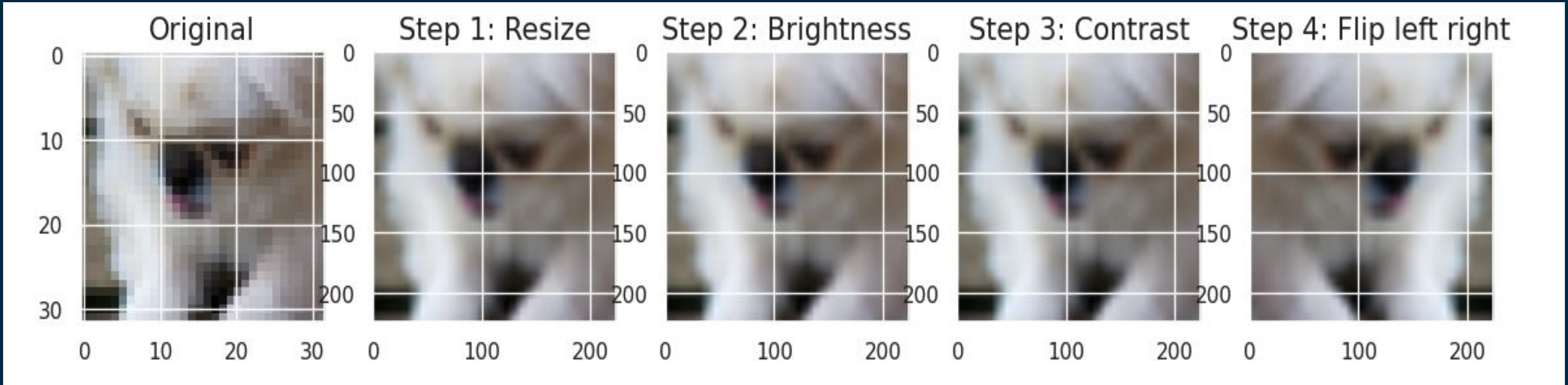


ResNet50 (87% test accuracy)



- Both models pre-trained on ImageNet (1.2M 224x224 images, 1000 classes)
- **VGG16 (2014)**: *shrinks kernel fields/pixel strides* & *adds convolutional layers*
- **ResNet (2015**: *deep residual networks* to solve for *vanishing gradients problem*

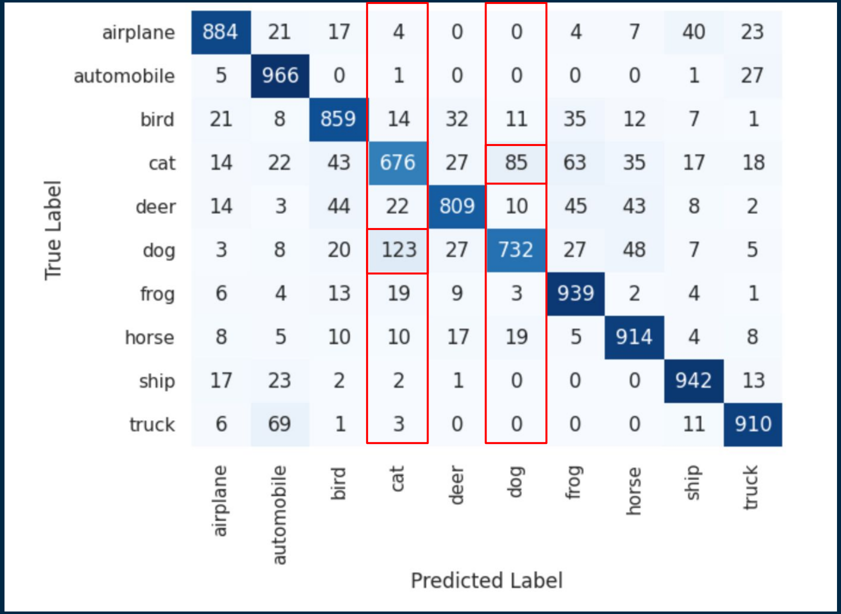
Pre-Processing & Data Augmentation Experiments



Original: 32 X 32 | **Resize: 64 x 64** | **Brightness: 0.3** | **Contrast: 2** | **Image Flip**

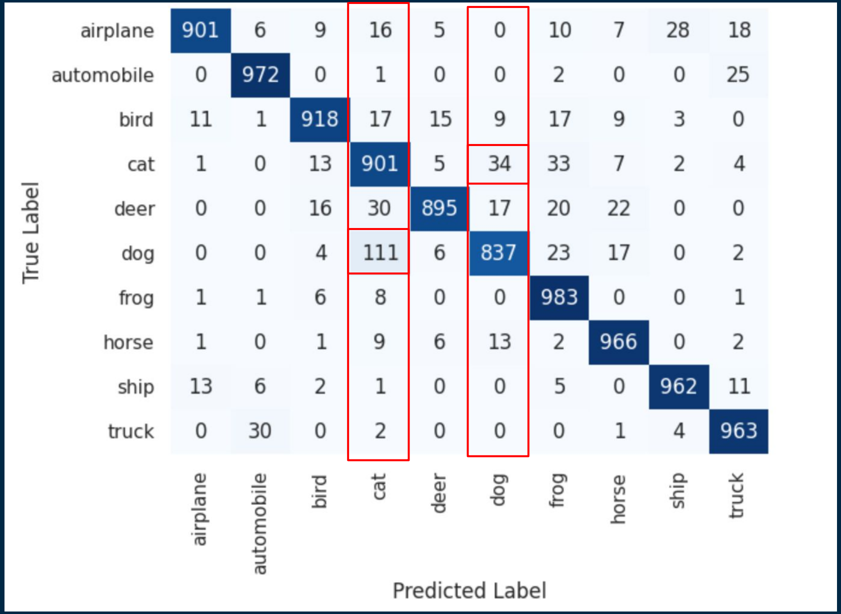
Preprocessing improves accuracy & precision, but incurs significant compute costs

VGG16 **Without** Pre Processing:



Testing Accuracy:84%
Animal Precision:81%

VGG16 **With** Pre Processing:



Testing Accuracy: 93% **(+9%)**
Animal Precision: 91% **(+10%)**

Pre-trained models & data pre-processing improves prediction accuracy, but our use case requires **smaller models, faster processing times, better class precision**

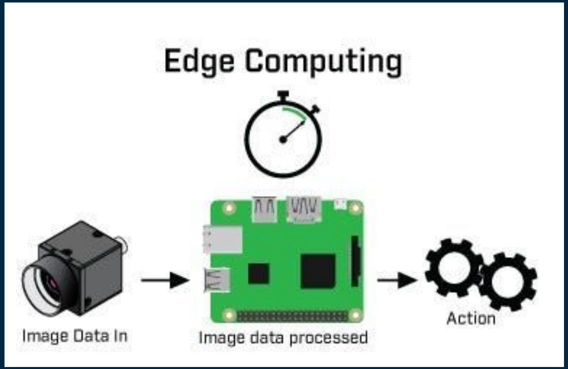
MODEL	TRAINING ACCURACY	VALIDATION ACCURACY	TEST ACCURACY	#LAYERS	ACTIVATION	OPTIMIZER	#PARAMETERS	PROCESSING TIME	ANIMAL PRECISION	NON-ANIMAL PRECISION
FFNN	.5377	.4936	.4931	1	relu	Adam	3,157,002	40s		
CNN1	.8765	.7118	.7037	8	relu	Adam	1,757,002	60s		
CNN2	.8484	.7450	.7339	13	relu	Adam	5,411,370	60s		
CNN3	.9423	.8428	.8366	23	relu	Adam	8,438,858	80s	.8197	.9343
CNN3 + Pre-Processing	.9756	.9107	.8341	23	relu	Adam	21,021,770	620s	.8087	.9335
VGG16 v4	.9943	.8612	.8563	16+7	relu	SGD	16,299,850	200s	.8337	.9388
VGG16 v4 + Pre-Processing	.9698	.9846	.9298	16+7	relu	SGD	17,872,714	2400s	.9213	.9642
ResNet50 v4	.9626	.8818	.8740	50+10	relu	Adam	26,745,738	1450s	.8626	.9490

Note: Data pre-processing steps includes resized training images to 64x64 resolution, adjusted image brightness (0.3), adjusted contrast factor (2), and randomly flipped training images.

Performance improvements of pre-trained models, data pre-processing not worth >10x training time required.

Summary Takeaways, Ethical Considerations & Limitations

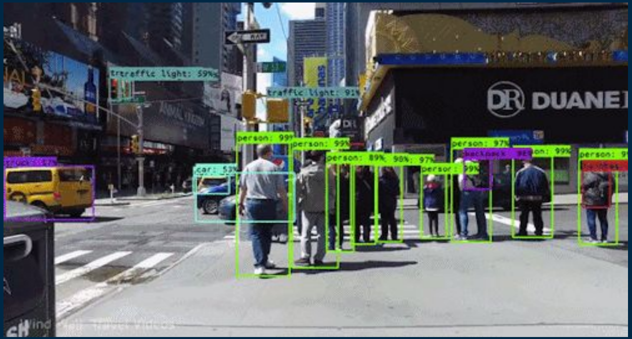
Summary Takeaway:
CNN v3 is optimal for our use case

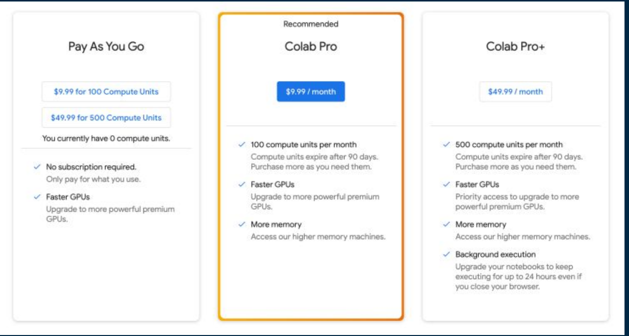


ML in edge devices like cameras requires:

1. Smaller models
2. Shorter processing times
3. Better precision between classes

Ethical Considerations & Limitations





Questions?