Image Generation and Particle Identification

NIKHEF, 25.04.2019 Fatih Bay

Previous/New Image Generation

For Particle Identification and Neutrino interaction classification: Image classification method used for DL is employed.

- Previous Method: Capture 2D Images from Event Display (time vs wire number)
- New Method: Get Space Point information for each event as txt file (X position vs Z position)

New method will provide more images in a shorter time and be helpful for another network studies through Image Classification

Event Capture vs Image Generation





Procedure: Capture images from 2D DUNE EVD and adjust its size as 224x224

417.914 600.774 92.9351 1 354.882 417.914 600.774 64.9557 354.627 417.514 600.937 69.6193 354.636 417.514 600.937 88.0338 354.222 417.514 601.471 108.841 .253 417.514 601.471 188.443 417.514 601.471 118.56 .031 417.115 601.841 83.2125 .061 417.115 601.841 66.1343 353.099 417.115 601.841 80.0305 353.845 417.115 601.841 92.4443 .859 417.115 601.841 73.0366 .125 416.715 602.004 36.4362 353.214 416.715 602.211 98.2747 352.872 416.715 602.744 192.751 352.901 416.715 602.744 77.9202 352.692 416.315 602.908 137.52 352.784 416.315 602.908 127.241 352.58 416.715 603.278 215.311 352.433 416.315 603.441 184.698 352.36 416.315 603.648 407.972 352.107 415.915 603.811 141.008 352.229 415.915 603.811 194.588 .564 415.915 604.345 356.863 .723 415.915 604.345 174.098 351.744 415.915 604.345 145.621 351.327 415.515 604.715 180.64 1 351.352 415.515 604.715 223.654 1 350.866 415.516 605.248 212.994 1 350.964 415.516 605.248 166.14 1 350.422 415.116 605.618 206.386



X-Z view

Procedure: Get space point information as X, Y, Z position associated with Hit Integral info in a txt file and then generate 224x224 images

5K AUTO 224D

- Dataset: 5000 events/images generated into 5 categories: e-, π⁰, p, μ⁻, π⁻ as TRAIN data with 4000 events and TEST data with 1000 events.
- 2D images were generated through txt files including position and Hit integral info.
- Image size is 224x224 pixels

1 354.855 417.914 600.774 92.9351 1 354.882 417.914 600.774 64.9557 354.627 417.514 600.937 69.6193 1 354.636 417.514 600.937 88.0338 354.222 417.514 601.471 108.841 354.253 417.514 601.471 188.443 1 354.305 417.514 601.471 118.56 1 353.031 417.115 601.841 83.2125 1 353.061 417.115 601.841 66.1343 1 353.099 417.115 601.841 80.0305 353.845 417.115 601.841 92.4443 353.859 417.115 601.841 73.0366 353.125 416.715 602.004 36.4362 353.214 416.715 602.211 98.2747 352.872 416.715 602.744 192.751 1 352.901 416.715 602.744 77.9202 352.692 416.315 602.908 137.52 1 352.784 416.315 602.908 127.241 1 352.58 416.715 603.278 215.311 1 352.433 416.315 603.441 184.698 1 352.36 416.315 603.648 407.972 352.107 415.915 603.811 141.008 352.229 415.915 603.811 194.588 351.564 415.915 604.345 356.863 351.723 415.915 604.345 174.098 351.744 415.915 604.345 145.621 351.327 415.515 604.715 180.64 351.352 415.515 604.715 223.654 350.866 415.516 605.248 212.994 350.964 415.516 605.248 166.14 1 350.422 415.116 605.618 206.386



5K AUTO Results

ResNet50



Accuracy: Correctly found image fraction

Loss: a value about how your prediction is well.

1 epoch: a process passing over all data (train+test) once

The results are preliminary, and the best trained model will be found by optimizing the process like increasing epoch, changing batch size, more data preparation, etc.

5K AUTO 224D Results over Confusion Matrix



	Precision	Recall	F1-score
e⁻	0.64	0.89	0.75
µ⁻	0.87	0.68	0.76
π-	0.42	0.80	0.55
π ⁰	0.78	0.39	0.52
р	0.69	0.29	0.41

Preliminary results

5 particles classification performance



Diagonal Entries: Correct prediction fraction Off-Diagonal Entries: Misidentification fraction Preliminary results

Particles	Correct Fraction [%]	Mis-Id [%]
e⁻	89	6 (π ⁰)
µ⁻	68	26 (π ⁻)
π-	80	7 (p)
π ⁰	39	46 (e⁻)
р	30	65 (π ⁻)

Comparison of Results

Particles	Correct Fraction [%]	Mis-Id [%]
e⁻	70	28.5 (π ⁰)
μ-	77.5	11 (π-)
π-	86	9.5 (p)
π ⁰	73	18 (e-)
р	64.5	27.5 (π ⁻)

5k new Image Generation (ResNet50)

Particles	Correct Fraction [%]	Mis-Id [%]	
e⁻	89	6 (π ⁰)	
μ-	68	26 (π ⁻)	
π-	80	7 (p)	
π ⁰	39	46 (e⁻)	
р	30	65 (π ⁻)	

140k images.		Classified Particle Type				
	Image, Network	e ⁻ [%]	γ[%]	μ- [%]	π ⁻ [%]	proton [%]
	HiRes, AlexNet	73.6 ± 0.7	81.3 ± 0.6	84.8 ± 0.6	73.1 ± 0.7	87.2 ± 0.5
	LoRes, AlexNet	64.1 ± 0.8	77.3 ± 0.7	75.2 ± 0.7	74.2 ± 0.7	85.8 ± 0.6
	HiRes, GoogLeNet	77.8 ± 0.7	83.4 ± 0.6	89.7 ± 0.5	71.0 ± 0.7	91.2 ± 0.5
:	LoRes, GoogLeNet	74.0 ± 0.7	74.0 ± 0.7	84.1 ± 0.6	75.2 ± 0.7	84.6 ± 0.6

MicroBooNE Results

https://arxiv.org/pdf/1611.05531.pdf

ν_e Neutrino Interaction Image Generation



Next

- Image Generation method is good enough to create dataset in a short time
- Optimize the images for Particle Identification and Neutrino event classification
 - OpenCV for image scaling
 - Use ADC values for generating images
 - Try space point size cut to reject bad images?
- Generate Neutrino Events

OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library https://opencv.org/about/

What is Confusion Matrix?

- It is a table used for checking classification performance in Machine Learning (Supervised learning).
- More detailed statistical information than test and train accuracy is provided through Confusion Matrix.
- It finds where the model confuses while it classifies particles.

Confusion Matrix II

		Positive (1)	Negative (0)	Particle Classifi	ication Analogy
d Values	Positive (1)	True Positive (TP)	False Positive (FP)	TP Π ⁰ This is pion-zero	Type I error (FP) C- This is pion-zero
Predicte	Negative (0)	False Negative (FN)	True Negative (TN)	Type II error (FN) Π ⁰ This is not pion-zero	TN E ⁻ This is not pion-zero

Actual Values

Performance measurement for particle classification!

Confusion Matrix III



Actual Values

 $Accuracy = \frac{TP + TN}{Total}$

Previous/New Image Generation II





Image generation 224x224









Previous/New Image Generation III

Event capture 224x224



15

5K 224D

2D DUNE EVD

- Dataset: 5000 events generated via larsoft and classified into 5 categories: e⁻, π⁰, p, μ⁻, π⁻ as TRAIN data with 4000 events and TEST data with 1000 events.
- 2D images were captured randomly by using standard event display of DUNE.
- Image size is 224x224 pixels



5K 224D Results



Loss: a value about how your prediction is well.

1 epoch: a process passing over all data (train+test) once

The results are preliminary, and the best trained model will be found by optimizing the process like increasing epoch, changing batch size, more data preparation, etc.

5K 224D Results over Confusion Matrix



5 particles classification performance



Preliminary results

Particles	Correct Fraction [%]	Mis-Id [%]
e⁻	70	28.5 (π ⁰)
µ⁻	77.5	11 (π ⁻)
π-	86	9.5 (p)
π ⁰	73	18 (e⁻)
р	64.5	27.5 (π ⁻)

Diagonal Entries: Correct prediction fraction Off-Diagonal Entries: Misidentification fraction