Time-of-Flight Vision Transformer #1

> why vision?

x	y	z	t	e	d
x_1	y_1	z_1	t_1	e_1	d_1
x_2	y_2	z_2	t_2	e_2	d_2
⋮					

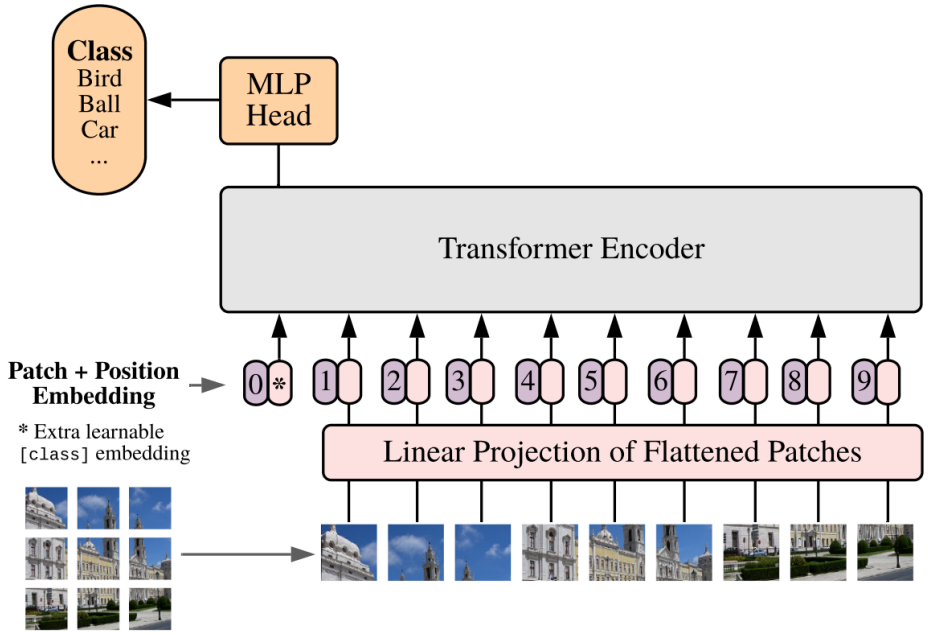
Convert to image

Image	Shower
Height [pixels]	n hits
Width [pixels]	1
Channels [RGB]	6 (x,y,z,t,e,d)

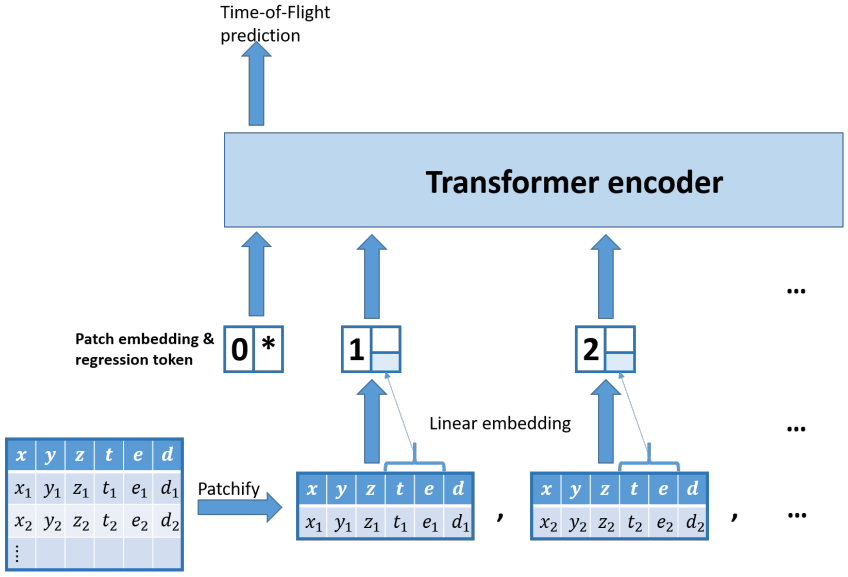
Input to ViT

Time-of-Flight Vision Transformer #2

’Normal’ Vision Transformer:

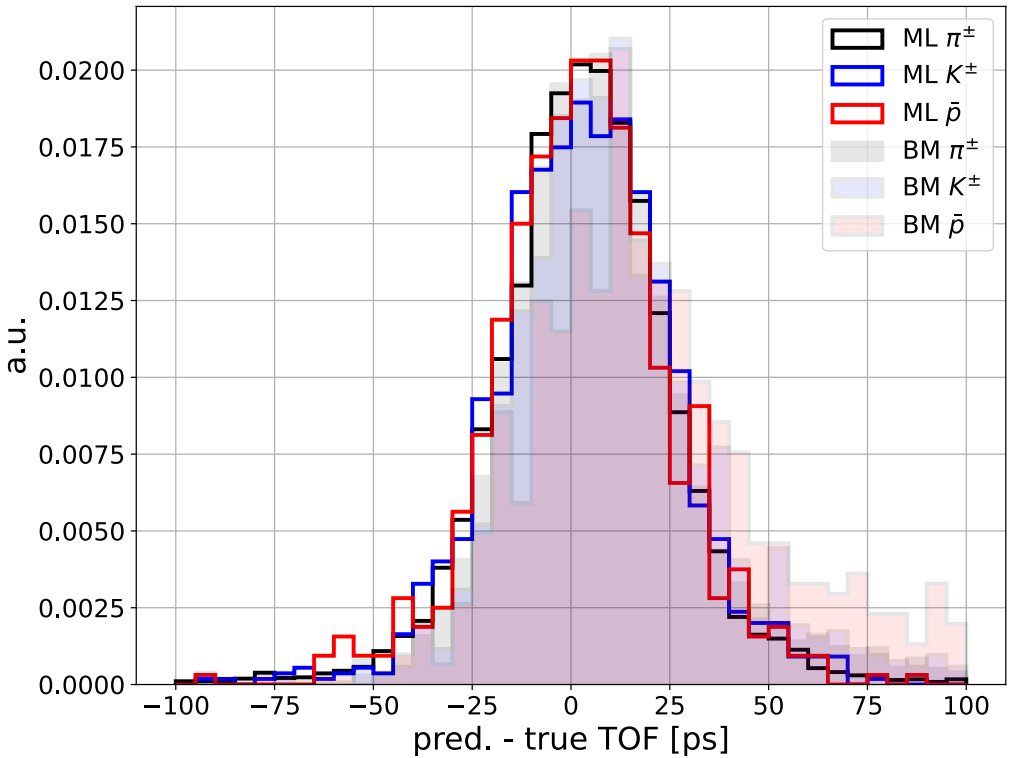


TOF Vision Transformer:



arXiv:2010.11929v2

TOFViT Results



- TOFViT beat the (improved) ILD benchmark
- TOFViT utilises 2 hits per layer
- only used hits from the first 10 layers
- smaller RMS90, less biased

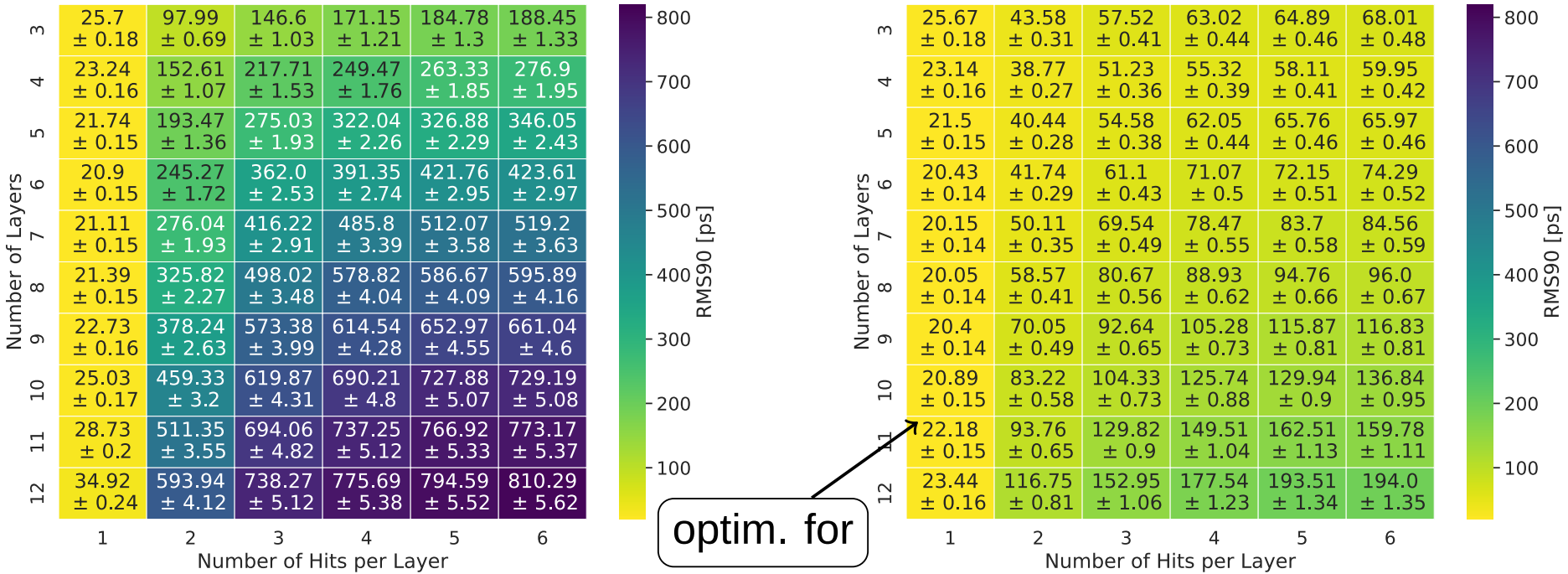
Network	RMS90	μ_{90}
Improved benchmark	$20.94 \pm ~~0.01~~$ 0.10	$9.77 \pm ~~0.02~~$ 0.20
TOFViT	$17.09 \pm ~~0.01~~$ 0.10	$3.23 \pm ~~0.02~~$ 0.20

Corrected!

Benchmark vs. Improved Benchmark

Benchmark Algorithm:

Improved Benchmark Algorithm:



*for Gaussians: RMS90 = $0.789 \cdot$RMS

A Closer Look at the Benchmark Algorithm

type	m [MeV]	%
$\pi^{\pm}$	140	85
$K^{\pm}$	490	10
$\overset{(-)}{p}$	940	5