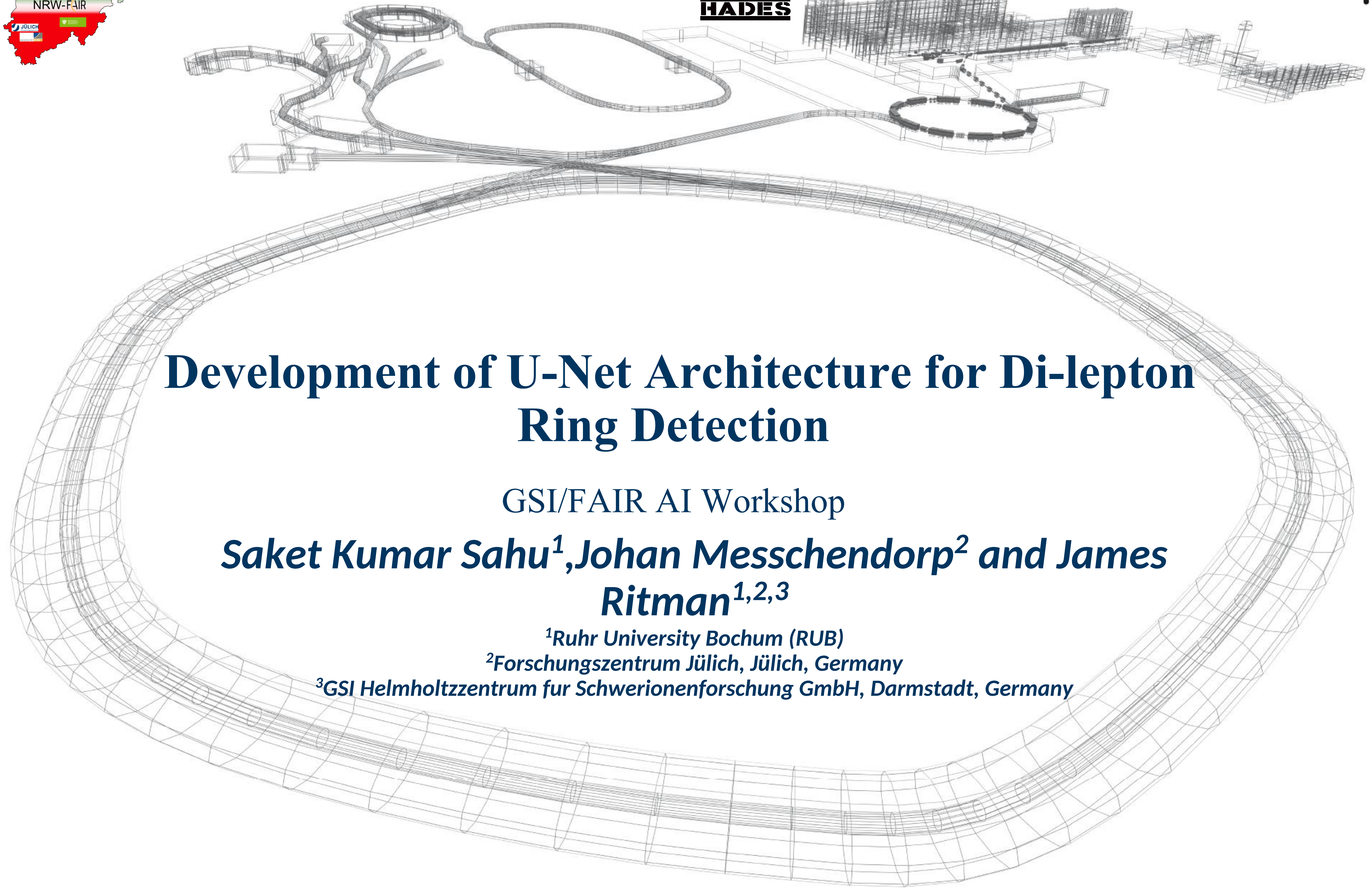


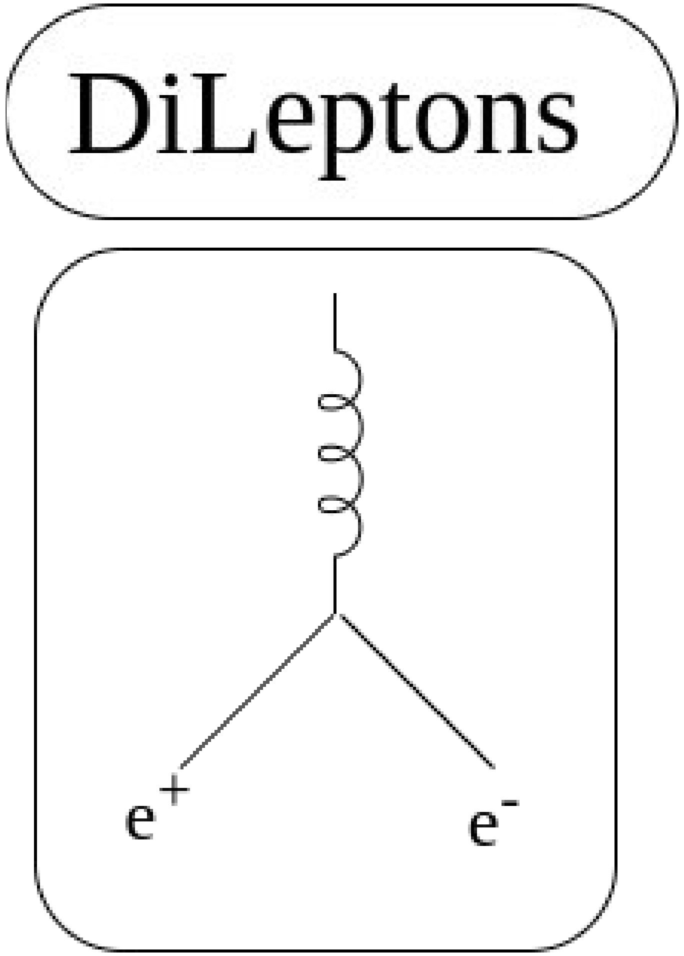
Development of U-Net Architecture for Di-lepton Ring Detection

GSI/FAIR AI Workshop

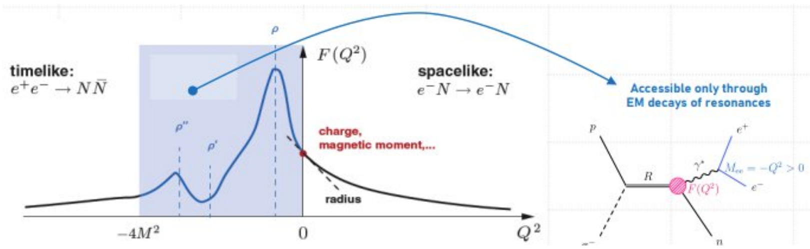
Saket Kumar Sahu[1],Johan Messchendorp[2] and James Ritman[1,2,3]

[1]Ruhr University Bochum (RUB)
[2]Forschungszentrum Jülich, Jülich, Germany
[3]GSI Helmholtzzentrum fur Schwerionenforschung GmbH, Darmstadt, Germany

Motivation



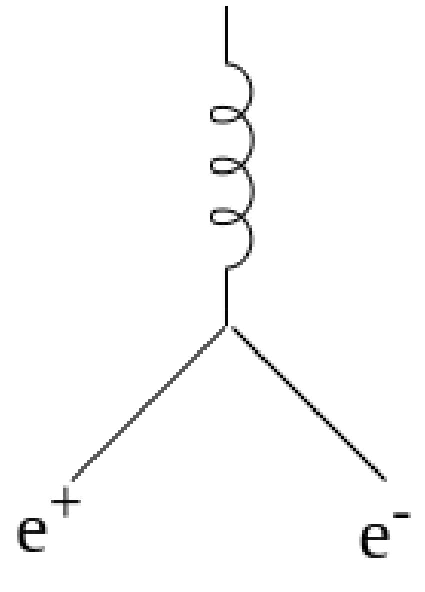
Motivation

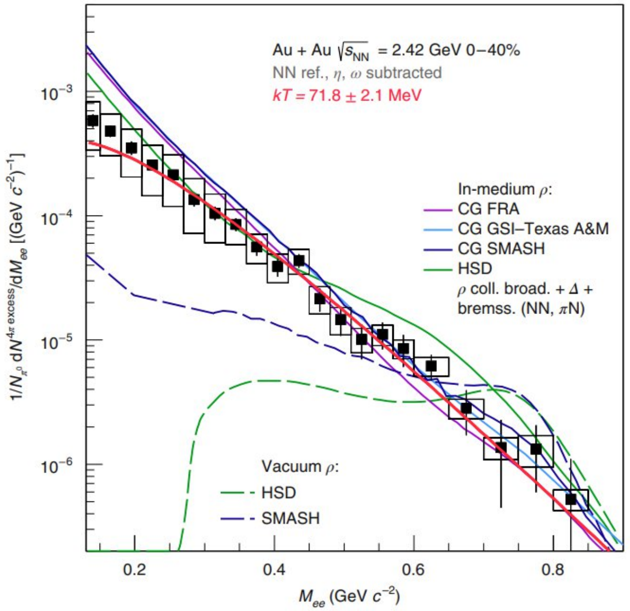


Baryon Structure through E.M. Form factors

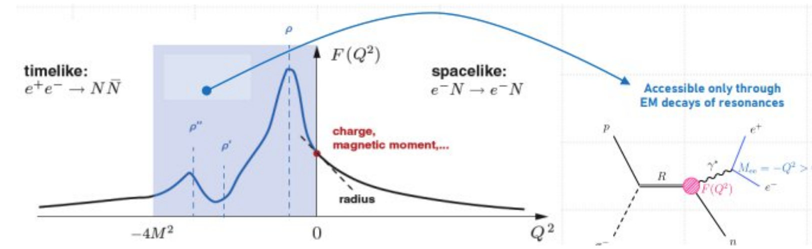
DiLeptons

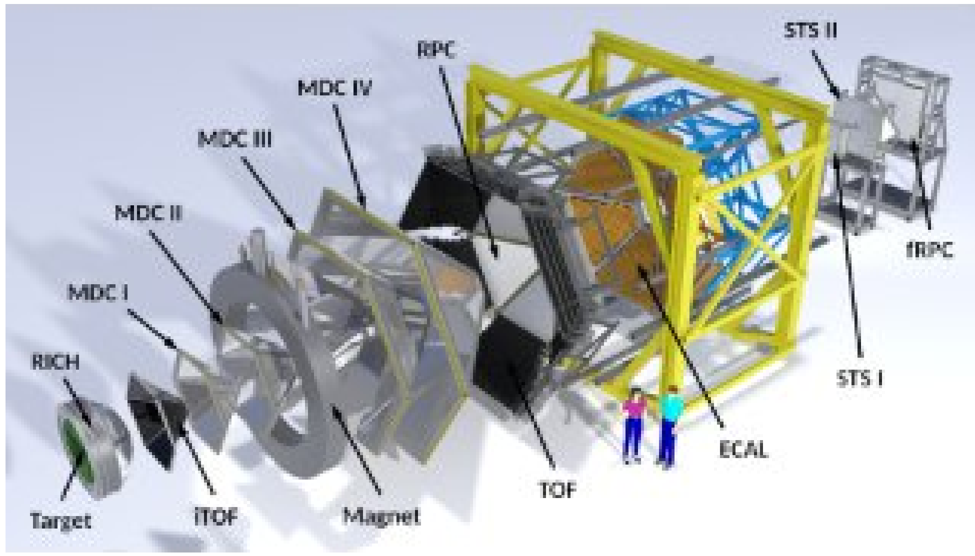
Production mechanisms axial and vector mesons





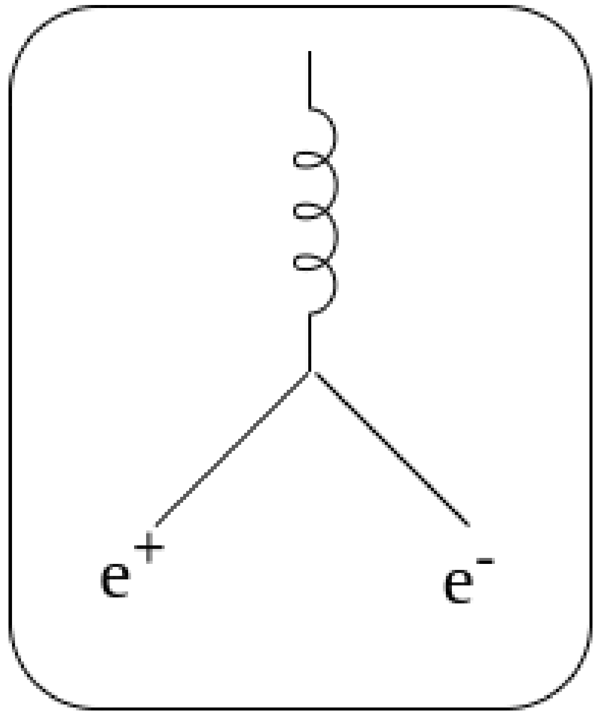
Motivation





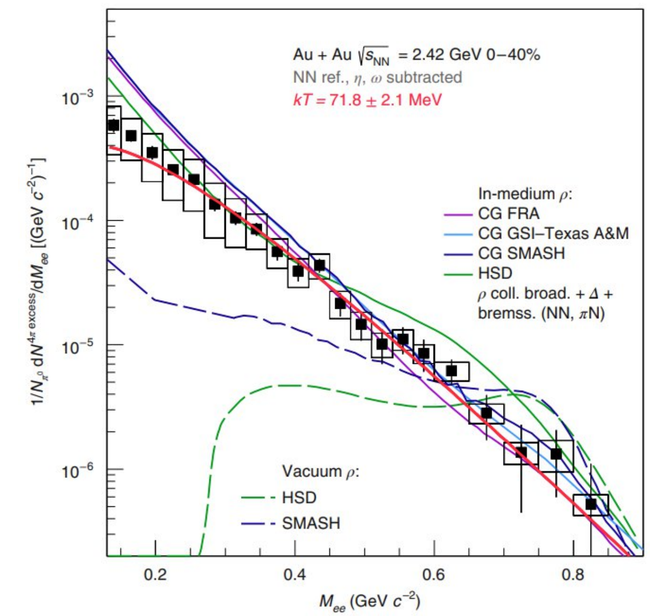
HADES-High Acceptance DiElectron Spectrometer

DiLeptons



Baryon Structure through E.M. Form factors

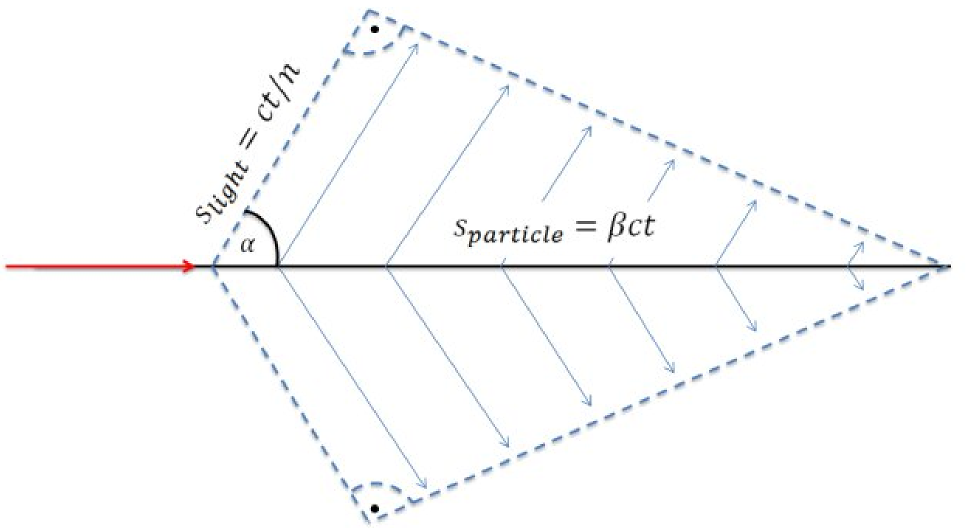
Production mechanisms axial and vector mesons



Cherenkov Radiation

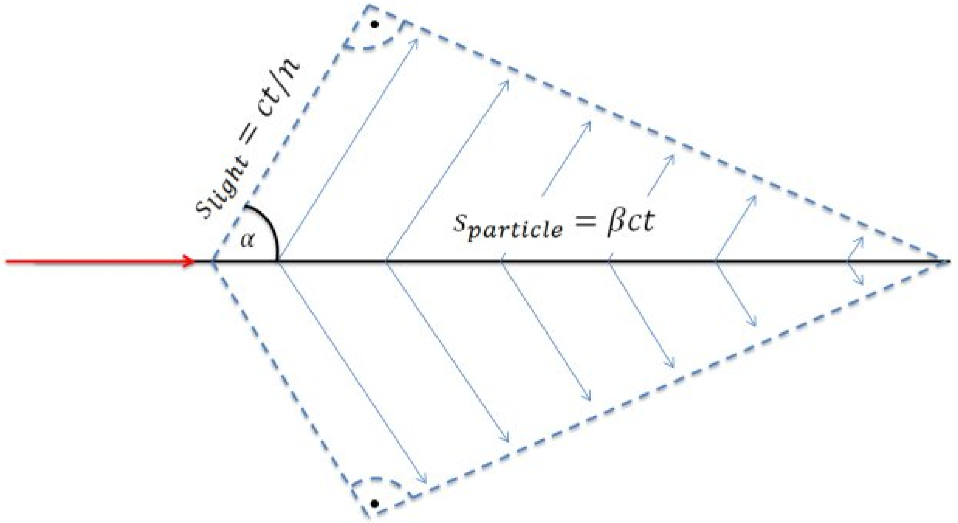
- **Cherenkov Radiation** (photons) emitted by a particle moving in a medium with a **speed greater** than the **speed of light in that medium**.

- These photons form a cone, which gives a **ring on the photodetector plane**.



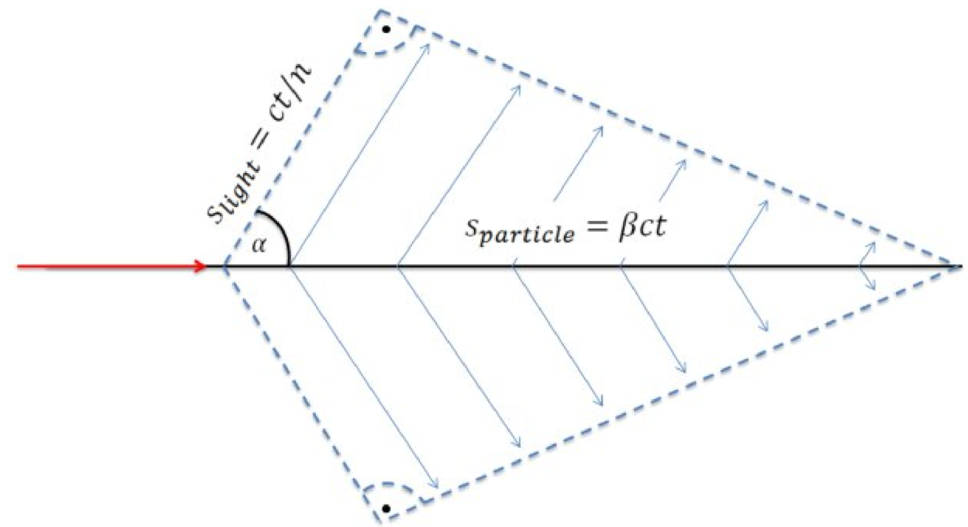
Cherenkov Radiation

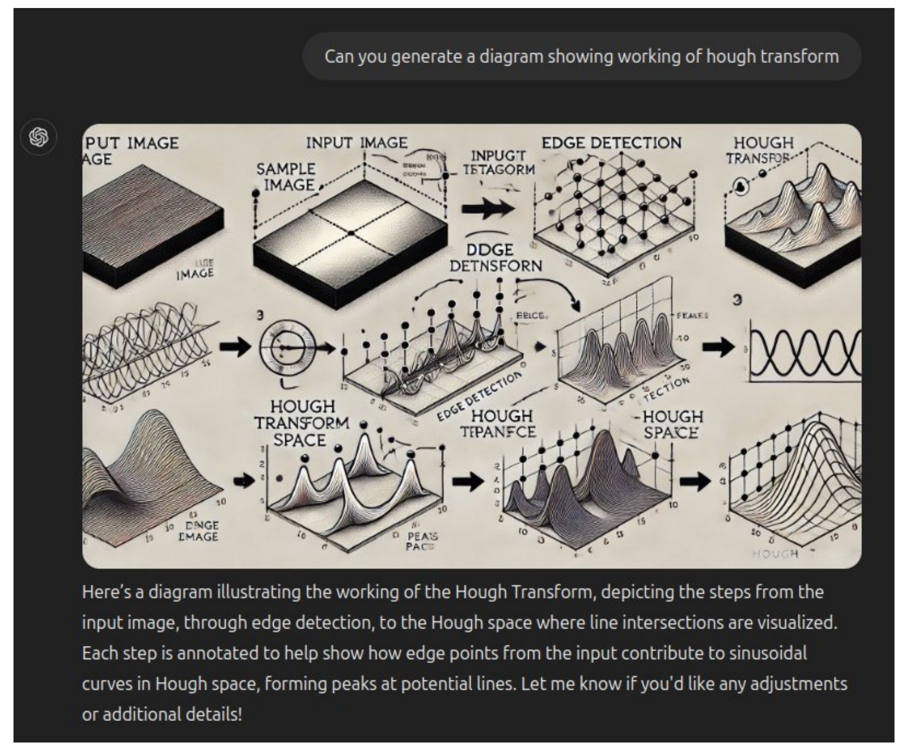
- **Cherenkov Radiation** (photons) emitted by a particle moving in a medium with a **speed greater** than the **speed of light in that medium**.
- These photons form a cone, which gives a **ring on the photodetector plane**.
- These rings are conventionally fitted with **Hough Transform**.



Cherenkov Radiation

- **Cherenkov Radiation** (photons) emitted by a particle moving in a medium with a **speed greater** than the **speed of light in that medium**.
- These photons form a cone, which gives a **ring on the photodetector plane**.
- These rings are conventionally fitted with **Hough Transform**.





Hough Transform

- The Hough Transform complexity **increases significantly** with the number of parameters involved i.e $O\,(A^{m-2})$ where A is the size of the image space and m is the number of parameters[1].

1. Shapiro, Linda and Stockman, George. "Computer Vision", Prentice-Hall, Inc. 2001

Hough Transform

- The Hough Transform complexity **increases significantly** with the number of parameters involved i.e $O\,(A^{m-2})$ where A is the size of the image space and m is the number of parameters[1].

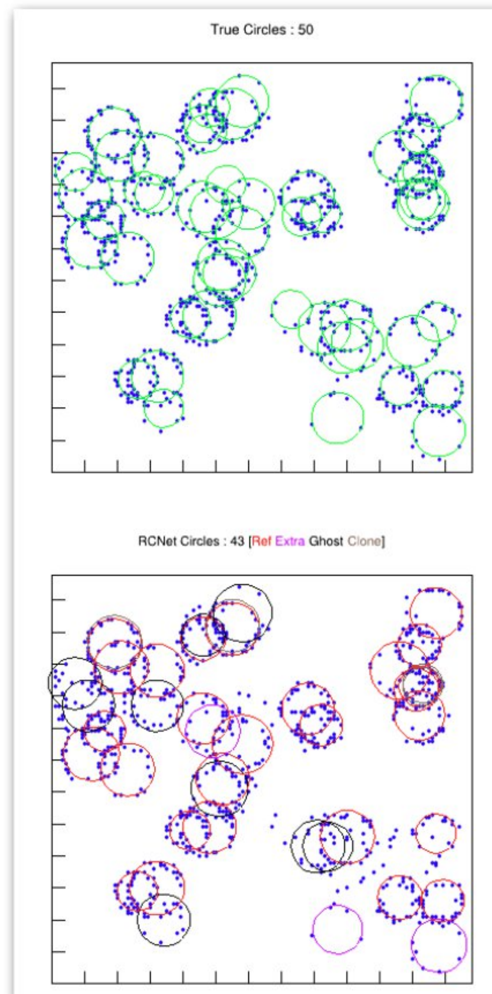
- **~10 MHz trigger rate** is expected in CBM/FAIR and a faster algorithm will be needed to speed up the **online analysis**.

- Can **Deep Learning algorithms** help?

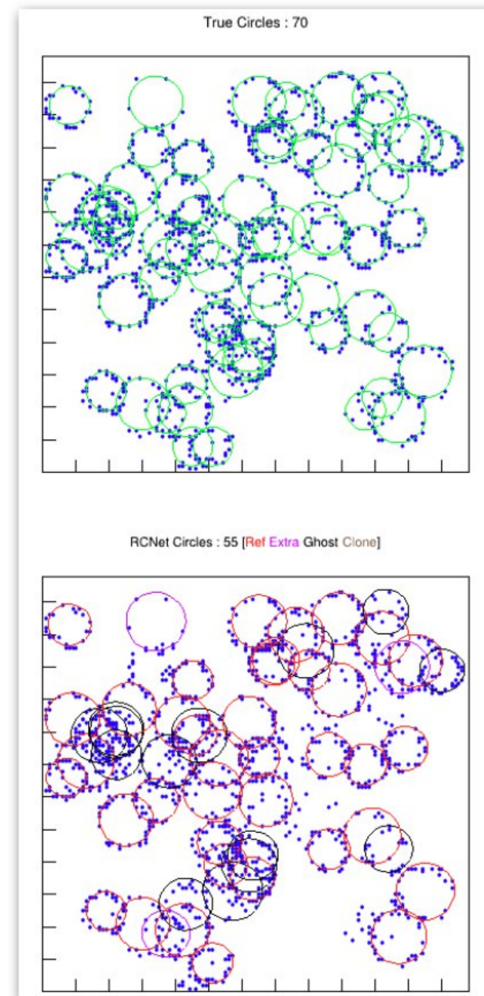
1. Shapiro, Linda and Stockman, George. "Computer Vision", Prentice-Hall, Inc. 2001

Hough Transform

- The Hough Transform complexity **increases significantly** with the number of parameters involved i.e $O\left(A^{m-2}\right)$ where A is the size of the image space and m is the number of parameters[1].
- **~10 MHz trigger rate** is expected in CBM/FAIR and a faster algorithm will be needed to speed up the **online analysis**.
- Can **Deep Learning algorithms** help?
- **U-Net** architecture already able to find rings in high density region[2].



50 Rings

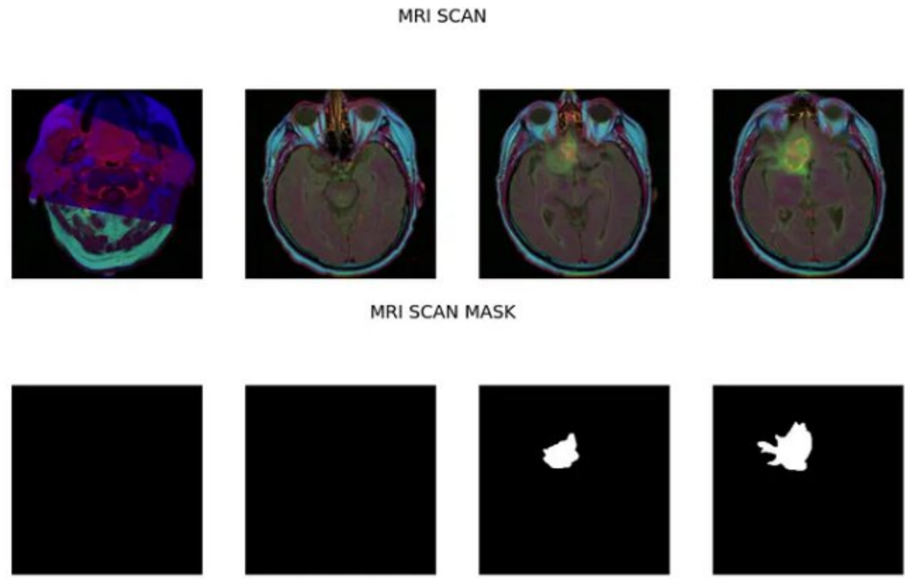


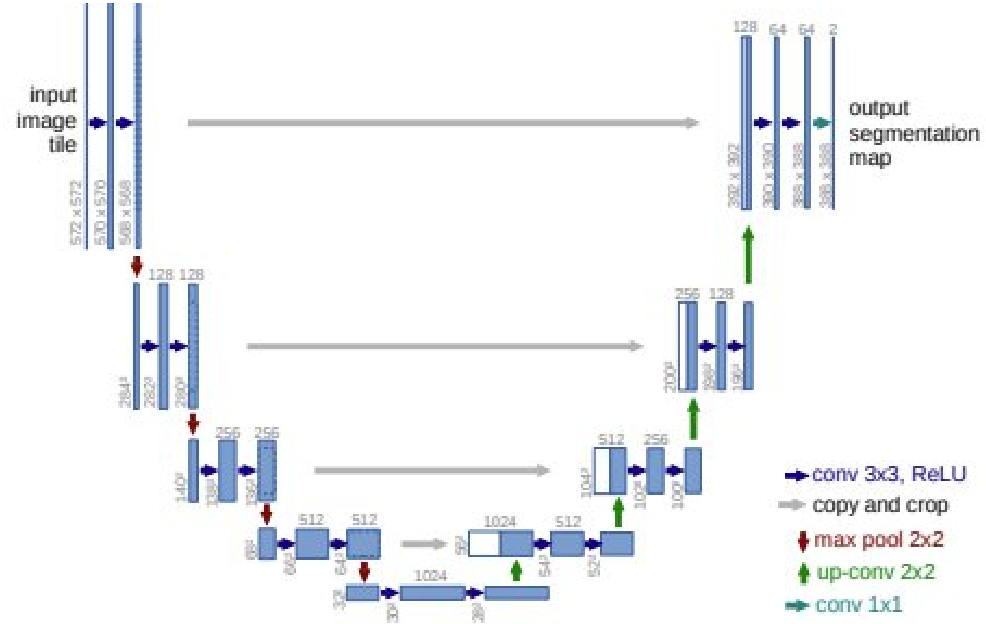
70 Rings

1. Shapiro, Linda and Stockman, George. "Computer Vision", Prentice-Hall, Inc. 2001
2. https://indico.nikhef.nl/event/4875/contributions/20267/attachments/8238/11744/Kisel_EuCAIFCon_2024.pdf

U-Net Model

- Originally proposed for **medical image segmentation** to search for tumors in MRI images[3].





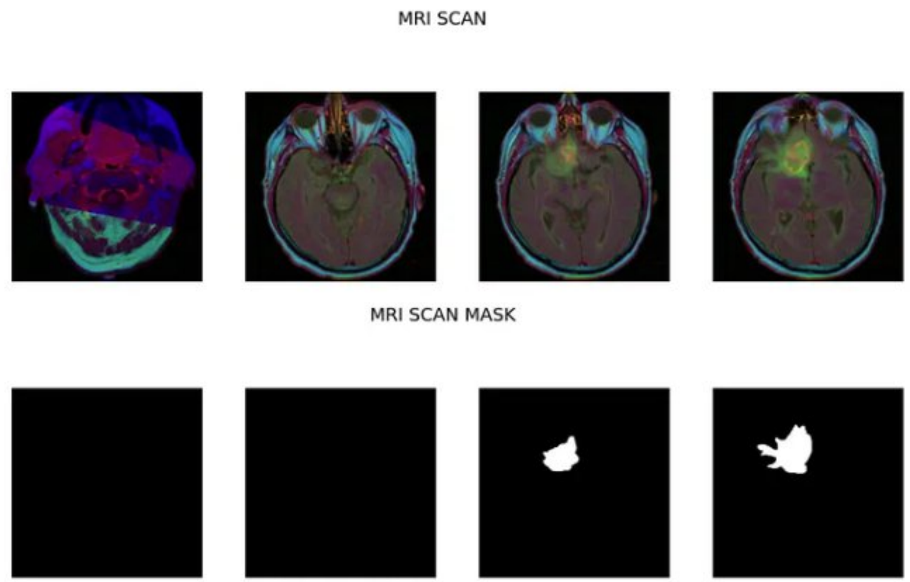
Unet architecture[4]

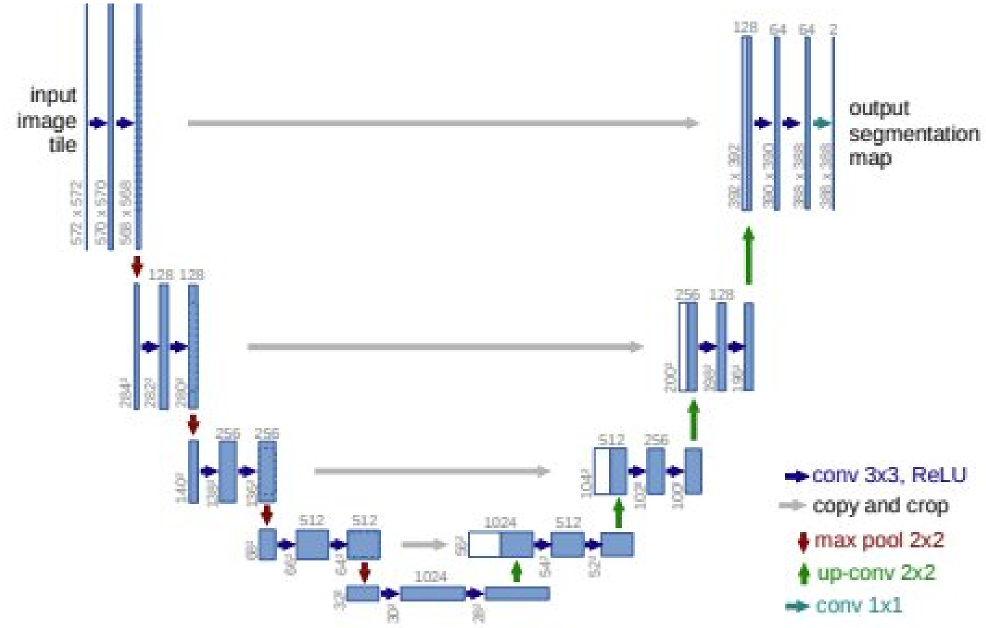
3. https://medium.com/@ashishjamarkattel123/image-segmentation-computer-vision-ea4e6f833bc5
4. https://arxiv.org/pdf/1505.04597.pdf

U-Net Model

- Originally proposed for **medical image segmentation** to search for tumors in MRI images[3].

- The U-Net architecture follows an **encoder-decoder cascade structure**, where the **encoder** gradually **compresses information** into a lower-dimensional representation and **decoder decodes** this information **back** to the original image dimension.





Unet architecture[4]

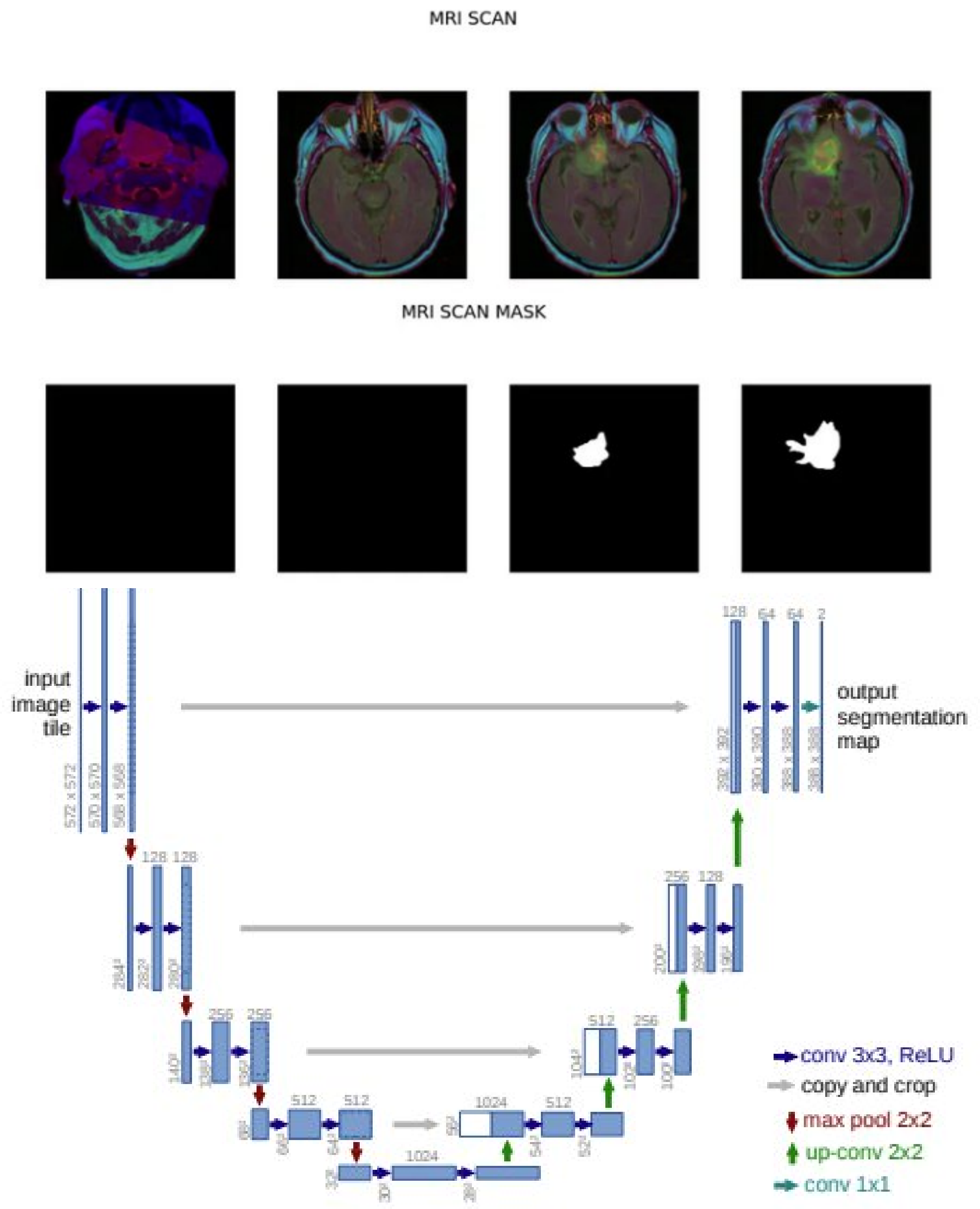
3. https://medium.com/@ashishjamarkattel123/image-segmentation-computer-vision-ea4e6f833bc5
4. https://arxiv.org/pdf/1505.04597.pdf

U-Net Model

- Originally proposed for **medical image segmentation** to search for tumors in MRI images[3].

- The U-Net architecture follows an **encoder-decoder cascade structure**, where the **encoder** gradually **compresses information** into a lower-dimensional representation and **decoder decodes** this information **back** to the original image dimension.

- The architecture gets an overall **U-shape**, which leads to the name **U-Net**.

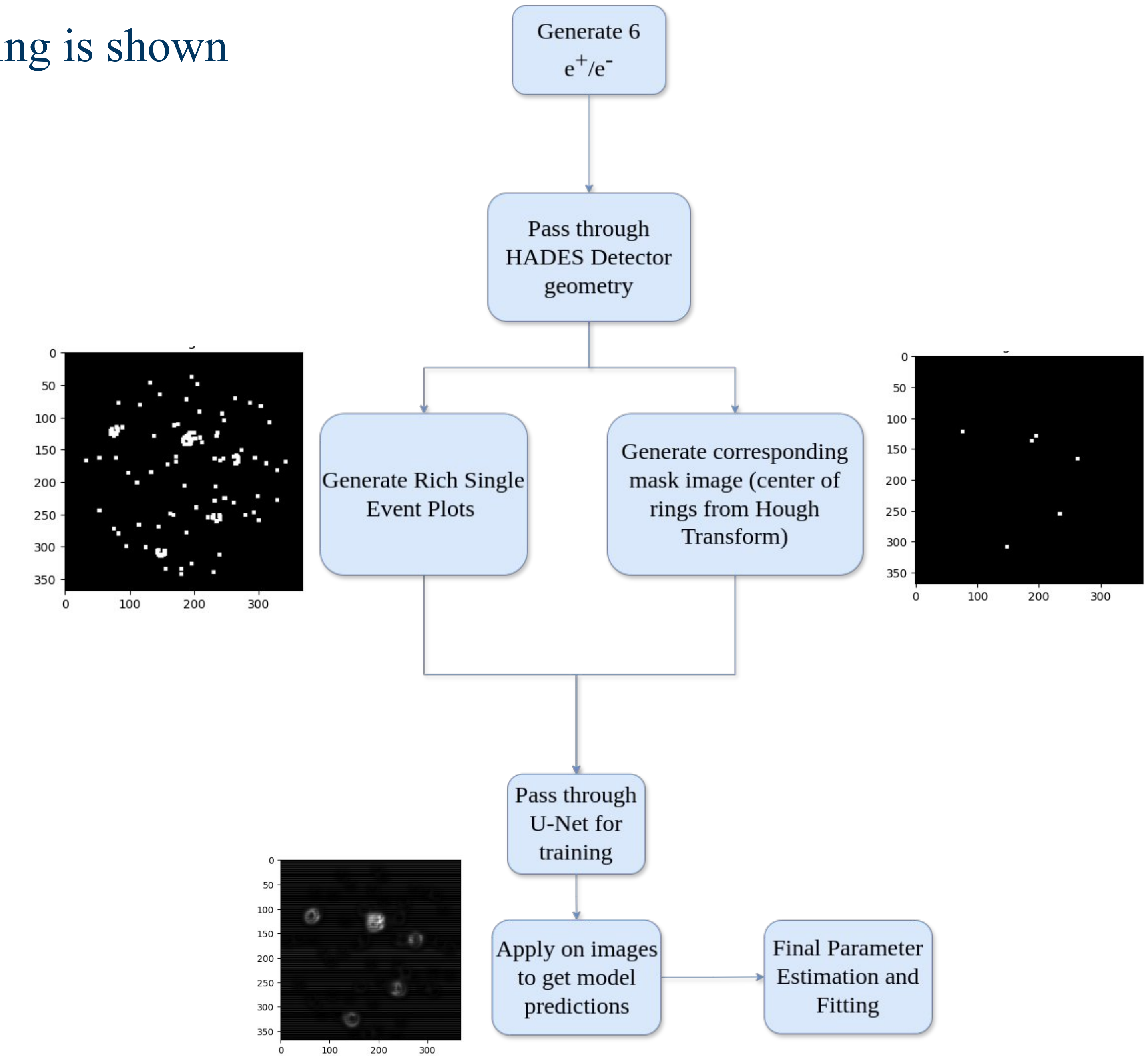


Unet architecture[4]

3. https://medium.com/@ashishjamarkattel123/image-segmentation-computer-vision-ea4e6f833bc5
4. https://arxiv.org/pdf/1505.04597.pdf

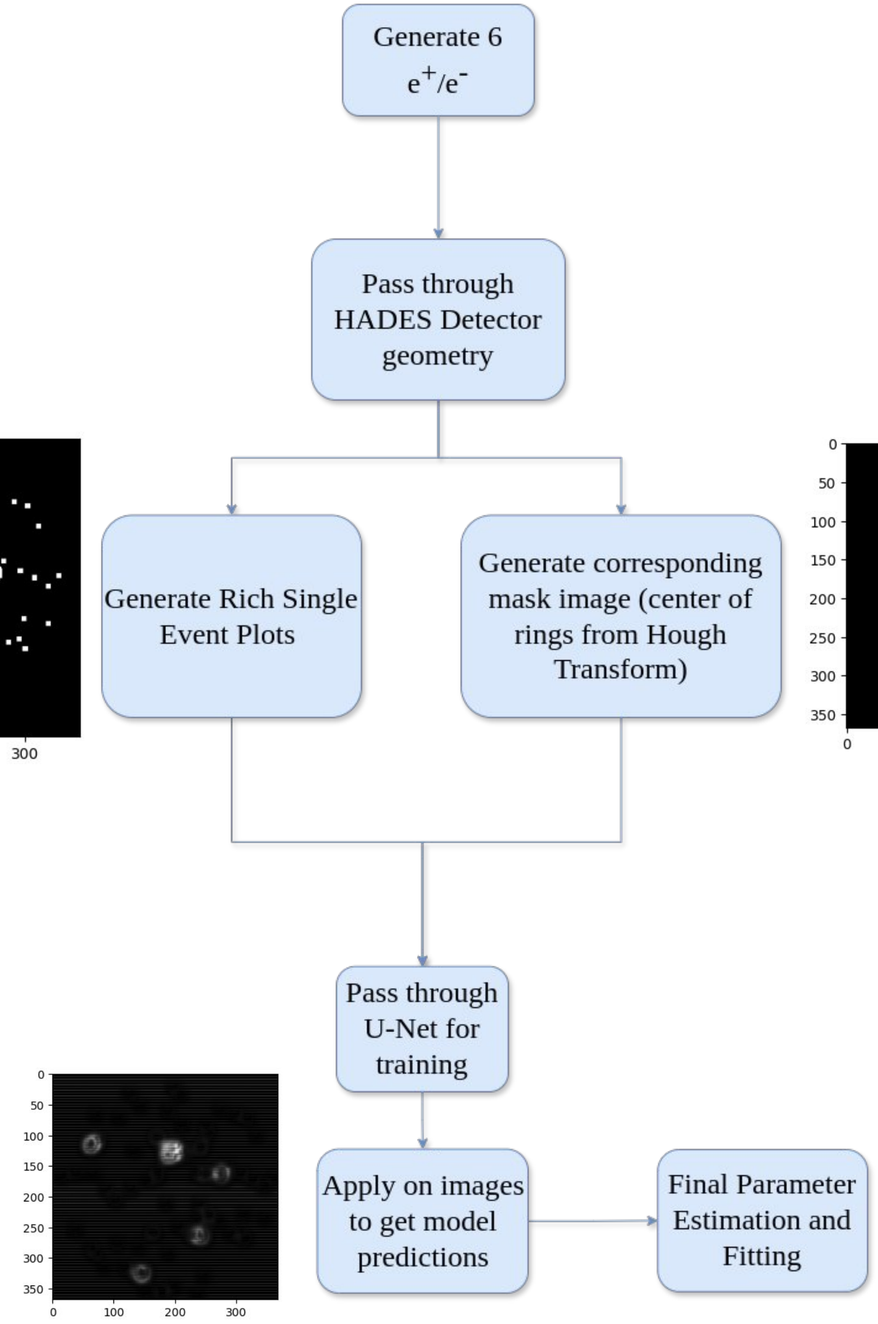
Training the Model

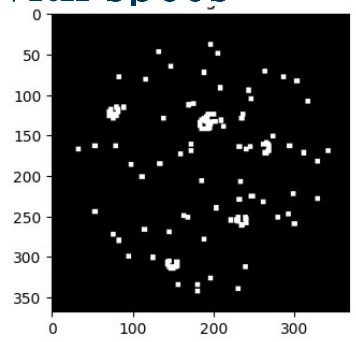
- The general work-flow followed for training is shown in the figure.

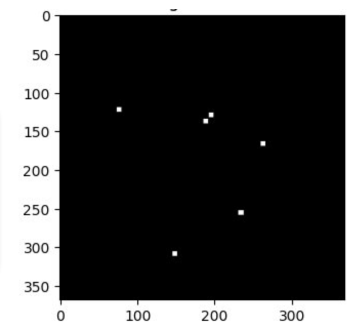


Training the Model

- The general work-flow followed for training is shown in the figure.
- The model is trained on 850 images.
- The time taken for training is ~5h in CPU with specs i7 (12 cores,32GB RAM) for 40 epochs.
- The loss function used is BCEWithLogitsLoss and Adam optimiser.







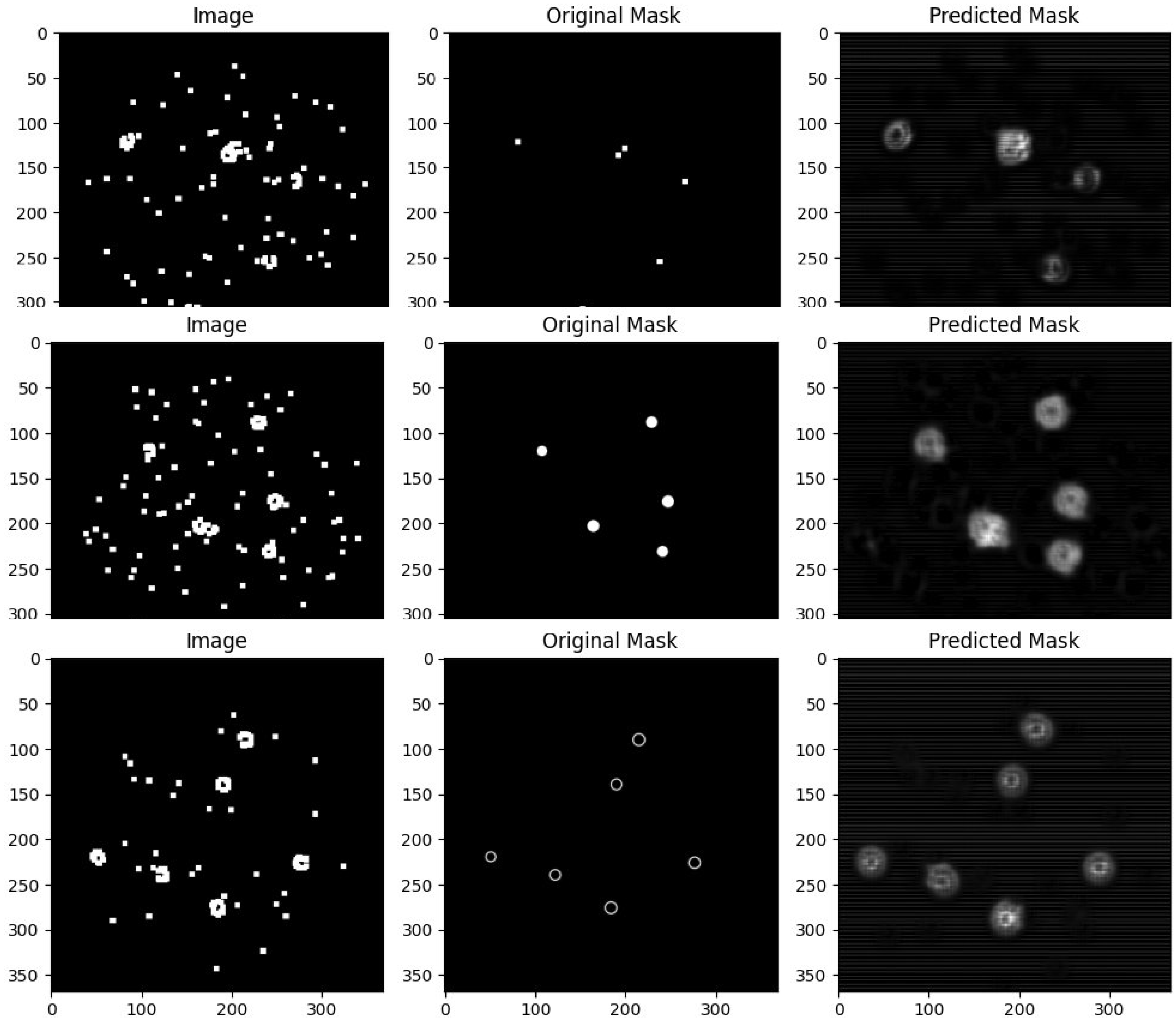
Model Predictions

- The **prediction time** of the **model** is ~ **0.04 sec (on CPU only)** compared to ~ **0.17 sec** taken **Hough Transform**.

Model Predictions

- The **prediction time** of the **model** is ~ **0.04 sec (on CPU only)** compared to ~ **0.17 sec** taken **Hough Transform**.

- Different kind of masks (filled circles, hollow circles and centres) were used for training.



Model Predictions

- A **threshold** (sigmoid function) on the model predictions was also applied.

Model Predictions

- A **threshold** (sigmoid function) on the model predictions was also applied.

- The optimal threshold was found by using **Dice coefficient**.

Model Predictions

- A **threshold** (sigmoid function) on the model predictions was also applied.
- The optimal threshold was found by using **Dice coefficient**.
- Dice coefficient is given by

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$

where $|X \cap Y|$ is the number of **overlapping pixel** between the predicted mask X and the ground truth mask Y,
$|X|$ is the number of pixels in the **predicted mask**,
$|Y|$ is the number of pixels in the **ground truth mask**.

Model Predictions

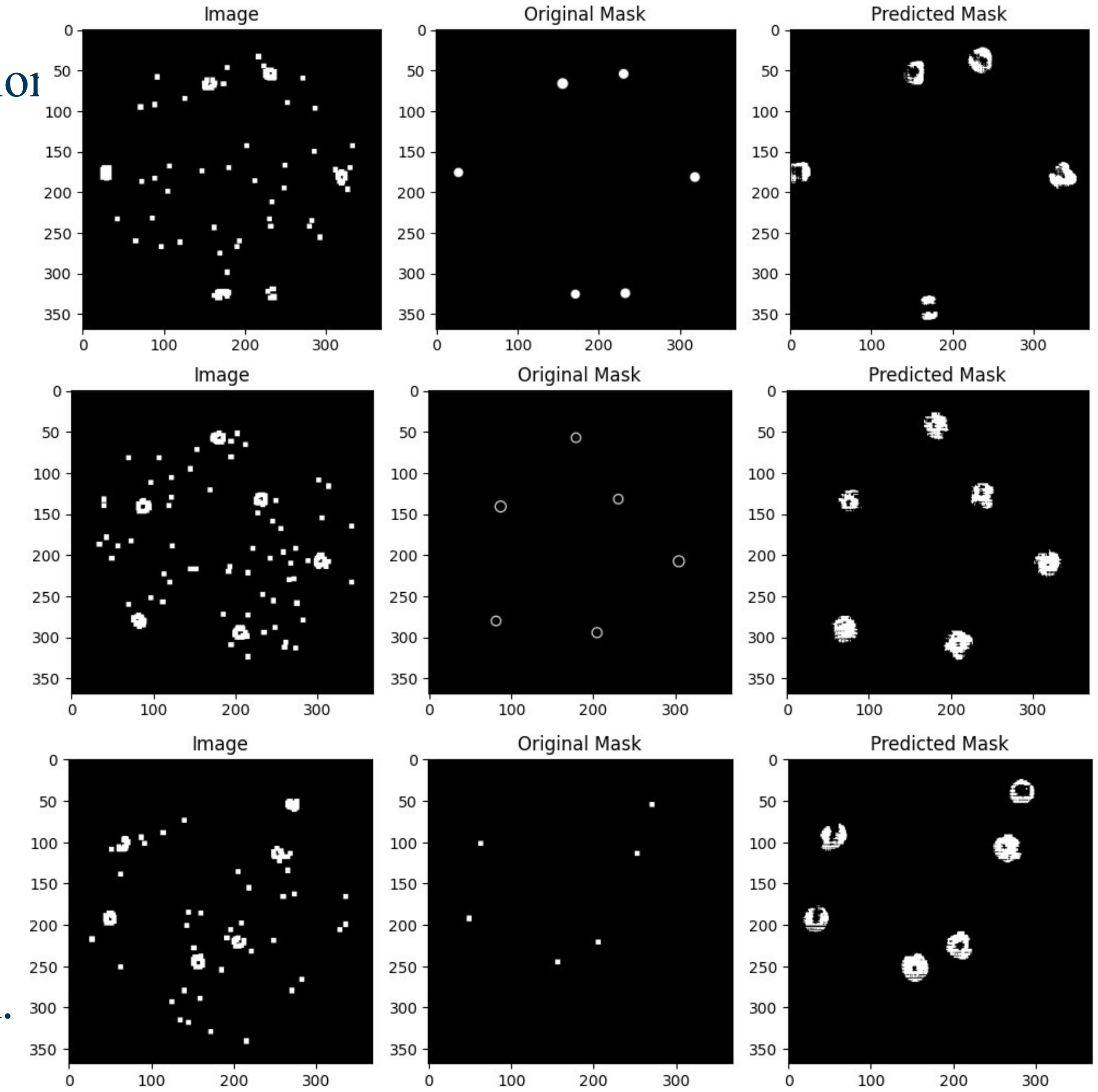
- A **threshold** (sigmoid function) on the model predictio was also applied.

- The optimal threshold was found by using **Dice coefficient**.

- Dice coefficient is given by

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$

where $|X \cap Y|$ is the number of **overlapping pixel** between the predicted mask X and the ground truth mask Y,
$|X|$ is the number of pixels in the **predicted mask**,
$|Y|$ is the number of pixels in the **ground truth mask**.



Conclusion and Outlook

- A preliminary U-Net architecture for ring detection has been developed.
- Improve the U-Net model training by including all kinds of masks, by applying augmentation techniques (like rotation,flipping, scaling, elastic transformations, etc) to make the model more robust to variations.
- Also work on hyperparameter optimisation.
- Implement the ring parameter extraction and ring fitting procedure.
- Implement the timing information of the pixel hits to training and extend the architecture to learn track information as well.
- Apply the model to low mass dileptons and high density ring regions and measure its performance.
- Export it to ONNX so it can be used for online implementation.

Conclusion and Outlook

- A preliminary U-Net architecture for ring detection has been developed.
- Improve the U-Net model training by including all kinds of masks, by applying augmentation techniques (like rotation,flipping, scaling, elastic transformations, etc) to make the model more robust to variations.
- Also work on hyperparameter optimisation.
- Implement the ring parameter extraction and ring fitting procedure.
- Implement the timing information of the pixel hits to training and extend the architecture to learn track information as well.
- Apply the model to low mass dileptons and high density ring regions and measure its performance.
- Export it to ONNX so it can be used for online implementation.

Δ Thank You Δ

Backup

.15
.27

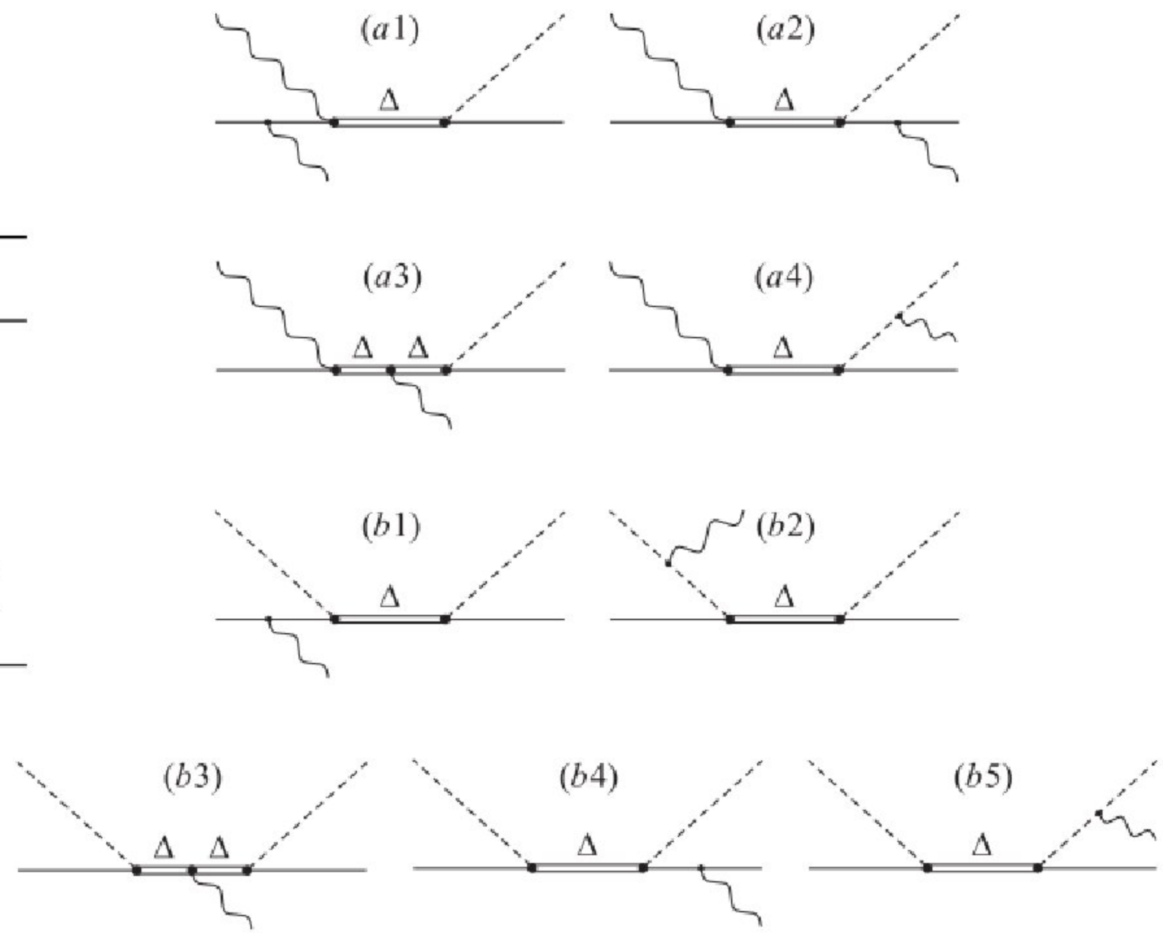


Fig. 2. Bremsstrahlung and Δ-resonant contributions to $N\pi\gamma'$ final states for pion photoproduction (a) and pion scattering (b). Only diagrams $(a3)$ and $(b3)$ are sensitive to the magnetic dipole moments μ_Δ.

Backup

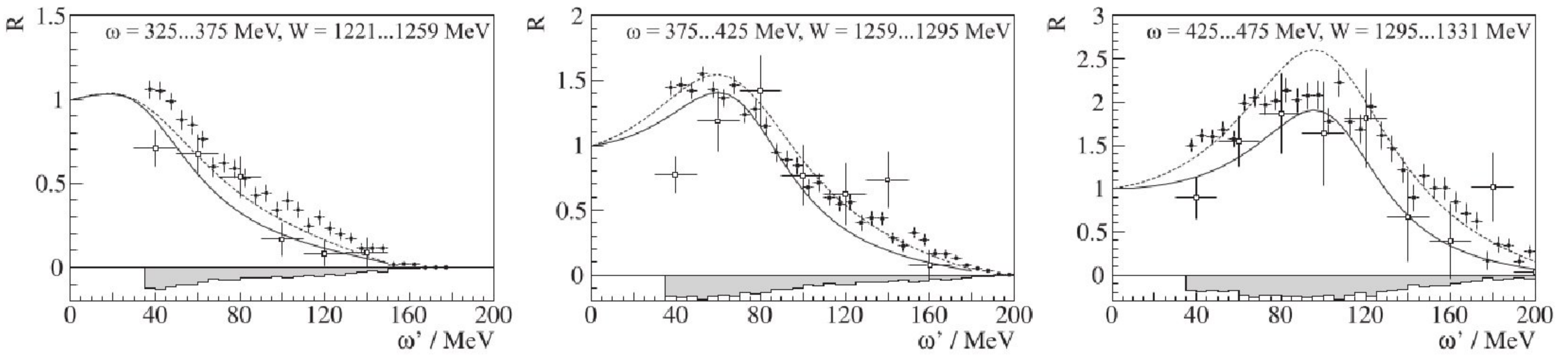


Fig. 18. Cross section ratio R at different ranges for beam energy ω and total c.m. energy W, respectively. Black points represent Crystal Ball / TAPS results, white squares are results from ref. [19]. Error bars denote statistical errors, grey shaded bands show absolute systematic uncertainties. Black lines are theoretical predictions (using $\kappa_{\Delta+} = 2.6$) of the unitary model from ref. [33] (dashed line) and the χEFT calculation from ref. [35] (solid line).

Backup

The unreduced (i.e. with `reduction` set to `'none'`) loss can be described as:

$$\ell(x, y) = L = \{l_1, \dots, l_N\}^\top, \quad l_n = -w_n \left[y_n \cdot \log \sigma(x_n) + (1 - y_n) \cdot \log(1 - \sigma(x_n)) \right],$$

where N is the batch size. If `reduction` is not `'none'` (default `'mean'`), then

$$\ell(x, y) = \begin{cases} \operatorname{mean}(L), & \text{if reduction} = \text{`mean';} \\ \operatorname{sum}(L), & \text{if reduction} = \text{`sum'.} \end{cases}$$

Backup

- The architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization.
- The contracting path contains encoder layers that capture contextual information and reduce the spatial resolution of the input, while the expansive path contains decoder layers that decode the encoded data and use the information from the contracting path via skip connections to generate a segmentation map.
- The network does not have any fully connected layers and only uses the valid part of each convolution, i.e., the segmentation map only contains the pixels, for which the full context is available in the input image.
- At the final layer a 1x1 convolution is used to map each 64 component feature vector to the desired number of classes.
- To allow a seamless tiling of the output segmentation map (see Figure 2), it is important to select the input tile size such that all 2x2 max-pooling operations are applied to a layer with an even x- and y-size.
- The skip connections from the contracting path are used to help the decoder layers locate and refine the features in the image.

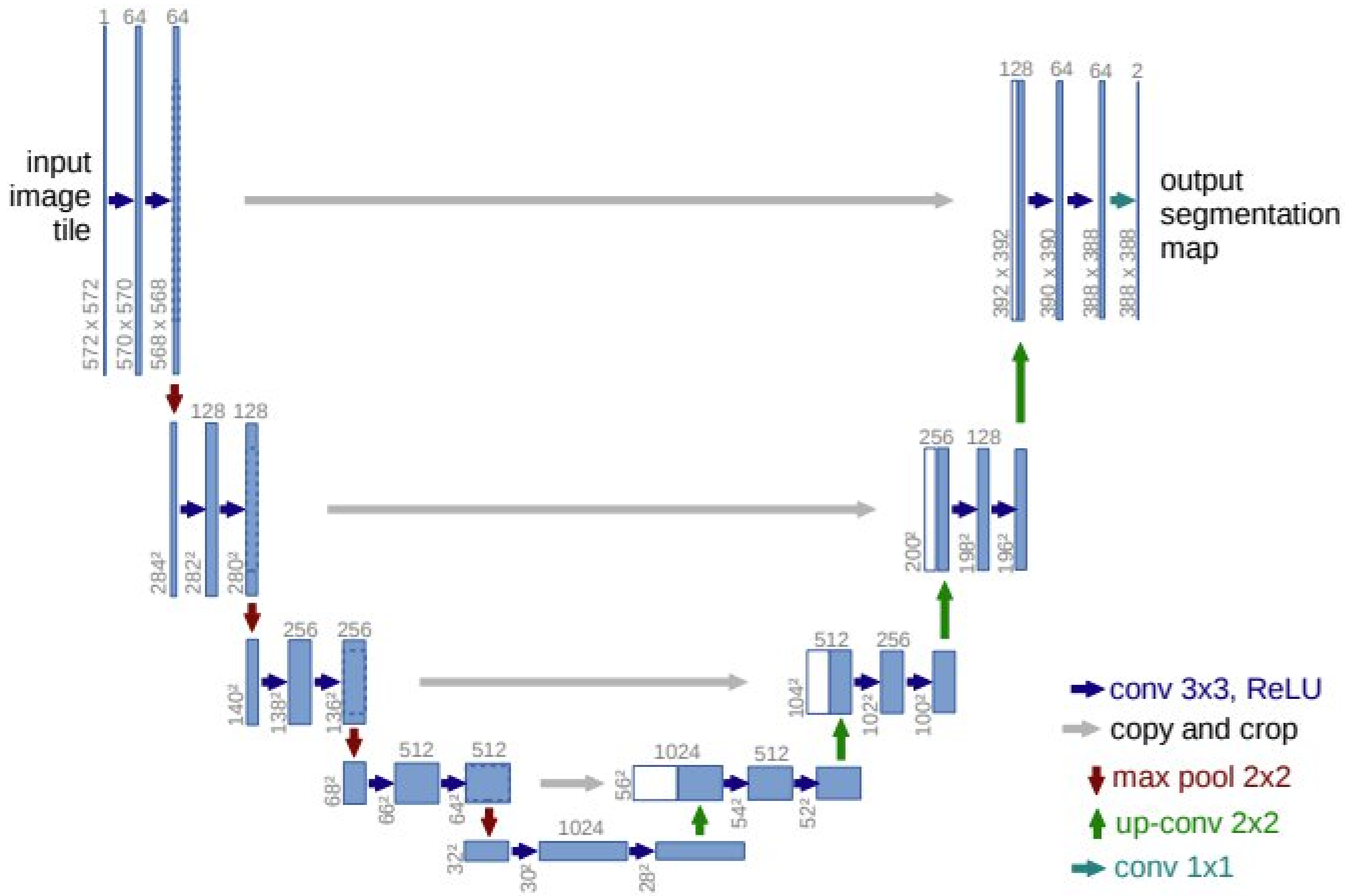
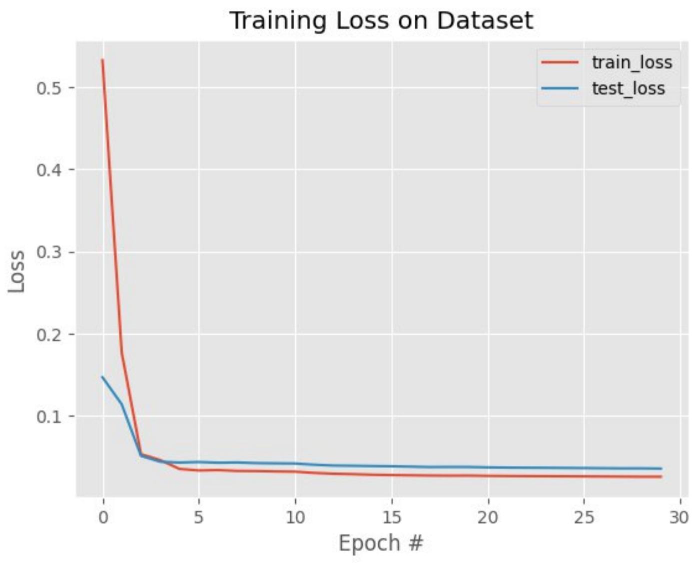
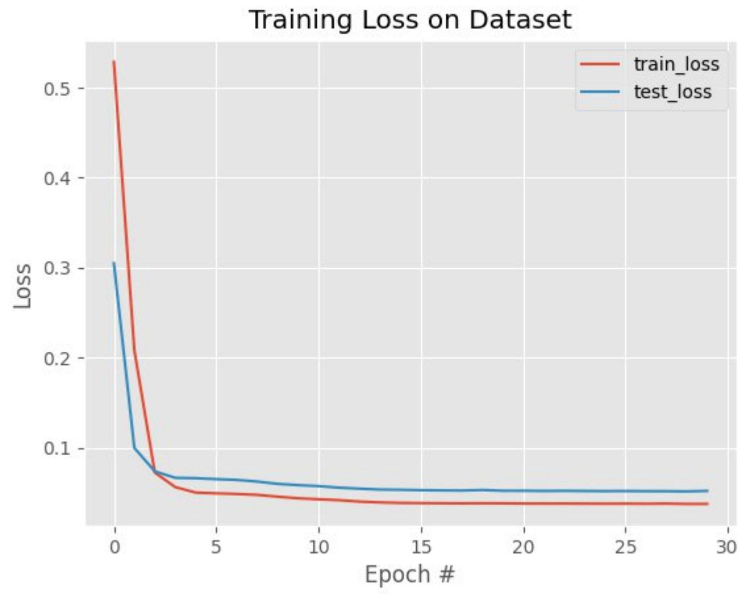
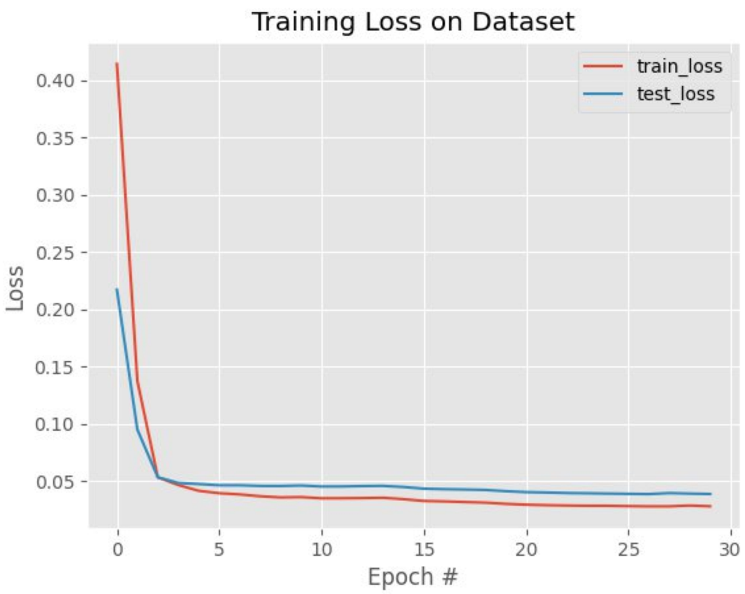


Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

Backup


Training Loss on Dataset
train_loss
test_loss
Loss
Epoch #


Training Loss on Dataset
train_loss
test_loss
Loss
Epoch #


Training Loss on Dataset
train_loss
test_loss
Loss
Epoch #