Deep Learning Methods for Jet Tagging and Process Classification Using Image Processing

J. G. Arneiro, Denise Moreira de Godoy, Alexandre Suaide
Instituto de Física da Universidade de São Paulo (IF-USP)
jhoaogabriel@usp.br

IFUSP Instituto de Física da USP

Universität Münster

Jet tagging

Introduction

What is the purpose of jet tagging?

What is the purpose of jet tagging?

- Jets are proxies to partons

What is the purpose of jet tagging?

- Jets are proxies to partons
- Understand where the jet comes from

What is the purpose of jet tagging?

- Jets are proxies to partons
- Understand where the jet comes from
- Reveal the physics behind its production[1]

[1] The ATLAS Collaboration, "Transforming jet flavour tagging at ATLAS", Nature Communications, vol. 17, p. 541, Jan. 2026

What is the purpose of jet tagging?

- Jets are proxies to partons
- Understand where the jet comes from
- Reveal the physics behind its production[1]
  - Precision measurements of the SM

[1] The ATLAS Collaboration, “Transforming jet flavour tagging at ATLAS”, Nature Communications, vol. 17, p. 541, Jan. 2026

What is the purpose of jet tagging?

- Jets are proxies to partons
- Understand where the jet comes from
- Reveal the physics behind its production[1]
    - Precision measurements of the SM
        - Heavy quarks

[1] The ATLAS Collaboration, “Transforming jet flavour tagging at ATLAS”, Nature Communications, vol. 17, p. 541, Jan. 2026

What is the purpose of jet tagging?

- Jets are proxies to partons
- Understand where the jet comes from
- Reveal the physics behind its production[1]
  - Precision measurements of the SM
    - Heavy quarks
    - Higgs boson

[1] The ATLAS Collaboration, “Transforming jet flavour tagging at ATLAS”, Nature Communications, vol. 17, p. 541, Jan. 2026

What is the purpose of jet tagging?

- Jets are proxies to partons
- Understand where the jet comes from
- Reveal the physics behind its production[1]
  - Precision measurements of the SM
    - Heavy quarks
    - Higgs boson
    - Top quark

[1] The ATLAS Collaboration, "Transforming jet flavour tagging at ATLAS", Nature Communications, vol. 17, p. 541, Jan. 2026

What is the purpose of jet tagging?

- Jets are proxies to partons
- Understand where the jet comes from
- Reveal the physics behind its production[1]
    - Precision measurements of the SM
        - Heavy quarks
        - Higgs boson
        - Top quark
        - Electroweak decays

[1] The ATLAS Collaboration, "Transforming jet flavour tagging at ATLAS", Nature Communications, vol. 17, p. 541, Jan. 2026

What is the purpose of jet tagging?

- Jets are proxies to partons
- Understand where the jet comes from
- Reveal the physics behind its production[1]
  - Precision measurements of the SM
    - Heavy quarks
    - Higgs boson
    - Top quark
    - Electroweak decays
  - Probe medium properties in A+A collisions

[1] The ATLAS Collaboration, "Transforming jet flavour tagging at ATLAS", Nature Communications, vol. 17, p. 541, Jan. 2026

What is the purpose of jet tagging?

- Jets are proxies to partons
- Understand where the jet comes from
- Reveal the physics behind its production[1]
    - Precision measurements of the SM
        - Heavy quarks
        - Higgs boson
        - Top quark
        - Electroweak decays
    - Probe medium properties in A+A collisions
    - Searches for BSM physics

[1] The ATLAS Collaboration, “Transforming jet flavour tagging at ATLAS”, Nature Communications, vol. 17, p. 541, Jan. 2026

What is the purpose of jet tagging?

- Jets are proxies to partons
- Understand where the jet comes from
- Reveal the physics behind its production[1]
    - Precision measurements of the SM
        - Heavy quarks
        - Higgs boson
        - Top quark
        - Electroweak decays
    - Probe medium properties in A+A collisions
    - Searches for BSM physics
    - Background rejection

[1] The ATLAS Collaboration, "Transforming jet flavour tagging at ATLAS", Nature Communications, vol. 17, p. 541, Jan. 2026

Motivation and objectives

The issue

The issue

- Massive amounts of raw data generated by detectors (e.g. in Run 3 the ALICE Experiment was giving a readout of **3.5 TB/s**)[2]

[2] D. Rohr on behalf of the ALICE Collaboration (2022). "Overview and 2022 data taking the experience of ALICE online and offline processing in Run 3", [Talk presented at the Kruger 2022: Discovery Physics at the LHC]. Johannesburg, South Africa

The issue

- Massive amounts of raw data generated by detectors (e.g. in Run 3 the ALICE Experiment was giving a readout of **3.5 TB/s**)[2]
    - Demands a **fast skimming** of the data, especially in the **triggering** systems

[2] D. Rohr on behalf of the ALICE Collaboration (2022). “Overview and 2022 data taking the experience of ALICE online and offline processing in Run 3”, [Talk presented at the Kruger 2022: Discovery Physics at the LHC]. Johannesburg, South Africa

The issue

- Massive amounts of raw data generated by detectors (e.g. in Run 3 the ALICE Experiment was giving a readout of **3.5 TB/s**)[2]
    - Demands a **fast skimming** of the data, especially in the **triggering** systems
- How can we reduce data size and extract events of interest?

[2] D. Rohr on behalf of the ALICE Collaboration (2022). "Overview and 2022 data taking the experience of ALICE online and offline processing in Run 3", [Talk presented at the Kruger 2022: Discovery Physics at the LHC]. Johannesburg, South Africa

The issue

- Massive amounts of raw data generated by detectors (e.g. in Run 3 the ALICE Experiment was giving a readout of **3.5 TB/s**)[2]
    - Demands a **fast skimming** of the data, especially in the **triggering** systems
- How can we reduce data size and extract events of interest?
    - **Machine learning** perform **quick** predictions and are already used for triggering

[2] D. Rohr on behalf of the ALICE Collaboration (2022). "Overview and 2022 data taking the experience of ALICE online and offline processing in Run 3", [Talk presented at the Kruger 2022: Discovery Physics at the LHC]. Johannesburg, South Africa

The issue

- Massive amounts of raw data generated by detectors (e.g. in Run 3 the ALICE Experiment was giving a readout of **3.5 TB/s**)[2]
  - Demands a **fast skimming** of the data, especially in the **triggering** systems
- How can we reduce data size and extract events of interest?
  - **Machine learning** perform **quick** predictions and are already used for triggering
  - Reduce, or get rid of, steps that can slow the process down

[2] D. Rohr on behalf of the ALICE Collaboration (2022). "Overview and 2022 data taking the experience of ALICE online and offline processing in Run 3", [Talk presented at the Kruger 2022: Discovery Physics at the LHC]. Johannesburg, South Africa

The issue

- Massive amounts of raw data generated by detectors (e.g. in Run 3 the ALICE Experiment was giving a readout of **3.5 TB/s**)[2]
  - Demands a **fast skimming** of the data, especially in the **triggering** systems
- How can we reduce data size and extract events of interest?
  - **Machine learning** perform **quick** predictions and are already used for triggering
  - Reduce, or get rid of, steps that can slow the process down
    - Best case scenario: trigger directly from raw data

[2] D. Rohr on behalf of the ALICE Collaboration (2022). "Overview and 2022 data taking the experience of ALICE online and offline processing in Run 3", [Talk presented at the Kruger 2022: Discovery Physics at the LHC]. Johannesburg, South Africa

The issue

- Massive amounts of raw data generated by detectors (e.g. in Run 3 the ALICE Experiment was giving a readout of **3.5 TB/s**)[2]
  - Demands a **fast skimming** of the data, especially in the **triggering** systems
- How can we reduce data size and extract events of interest?
  - **Machine learning** perform **quick** predictions and are already used for triggering
  - Reduce, or get rid of, steps that can slow the process down
    - Best case scenario: trigger directly from raw data
- Mechanisms of data processing

[2] D. Rohr on behalf of the ALICE Collaboration (2022). "Overview and 2022 data taking the experience of ALICE online and offline processing in Run 3", [Talk presented at the Kruger 2022: Discovery Physics at the LHC]. Johannesburg, South Africa

The issue

- Massive amounts of raw data generated by detectors (e.g. in Run 3 the ALICE Experiment was giving a readout of **3.5 TB/s**)[2]
  - Demands a **fast skimming** of the data, especially in the **triggering** systems
- How can we reduce data size and extract events of interest?
  - **Machine learning** perform **quick** predictions and are already used for triggering
  - Reduce, or get rid of, steps that can slow the process down
    - Best case scenario: trigger directly from raw data
- Mechanisms of data processing
  - Track reconstruction

[2] D. Rohr on behalf of the ALICE Collaboration (2022). "Overview and 2022 data taking the experience of ALICE online and offline processing in Run 3", [Talk presented at the Kruger 2022: Discovery Physics at the LHC]. Johannesburg, South Africa

The issue

- Massive amounts of raw data generated by detectors (e.g. in Run 3 the ALICE Experiment was giving a readout of **3.5 TB/s**)[2]
  - Demands a **fast skimming** of the data, especially in the **triggering** systems
- How can we reduce data size and extract events of interest?
  - **Machine learning** perform **quick** predictions and are already used for triggering
  - Reduce, or get rid of, steps that can slow the process down
    - Best case scenario: trigger directly from raw data
- Mechanisms of data processing
  - Track reconstruction
  - Particle identification

[2] D. Rohr on behalf of the ALICE Collaboration (2022). "Overview and 2022 data taking the experience of ALICE online and offline processing in Run 3", [Talk presented at the Kruger 2022: Discovery Physics at the LHC]. Johannesburg, South Africa

The issue

- Massive amounts of raw data generated by detectors (e.g. in Run 3 the ALICE Experiment was giving a readout of **3.5 TB/s**)[2]
  - Demands a **fast skimming** of the data, especially in the **triggering** systems
- How can we reduce data size and extract events of interest?
  - **Machine learning** perform **quick** predictions and are already used for triggering
  - Reduce, or get rid of, steps that can slow the process down
    - Best case scenario: trigger directly from raw data
- Mechanisms of data processing
  - Track reconstruction
  - Particle identification
  - **Jet reconstruction**

[2] D. Rohr on behalf of the ALICE Collaboration (2022). "Overview and 2022 data taking the experience of ALICE online and offline processing in Run 3", [Talk presented at the Kruger 2022: Discovery Physics at the LHC]. Johannesburg, South Africa

Objectives

➔ Develop an **AI model** that can “sense” the **presence of jets** (and their flavor) in an event just by looking at **low level particle variables**

Objectives

- ➔ Develop an **AI model** that can “sense” the **presence of jets** (and their flavor) in an event just by looking at **low level particle variables**
  - ◆ **Without** jet **reconstruction**

Objectives

➔ Develop an **AI model** that can “sense” the **presence of jets** (and their flavor) in an event just by looking at **low level particle variables**

- **Without** jet **reconstruction**

Jet images

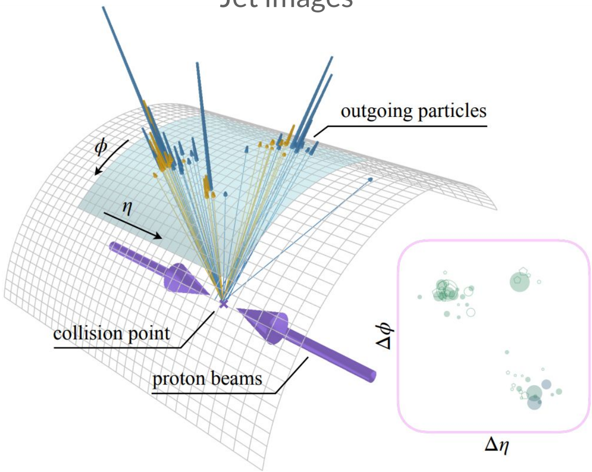


Image from [3]

[3] H. Qu, C. Li, and S. Qian, “Particle Transformer for Jet Tagging”, 2024.

Objectives

➔ Develop an **AI model** that can "sense" the **presence of jets** (and their flavor) in an event just by looking at **low level particle variables**

- **Without** jet **reconstruction**

Jet images

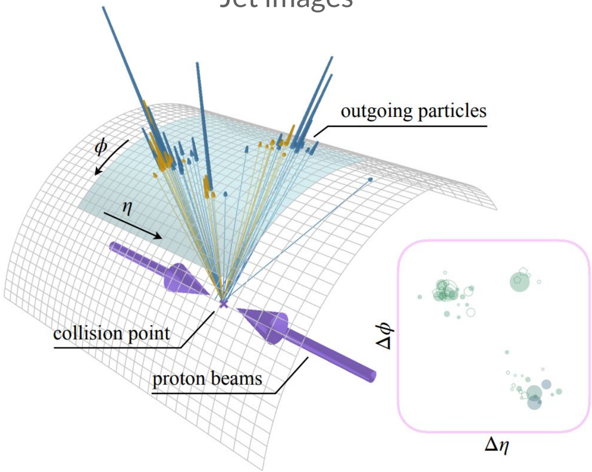
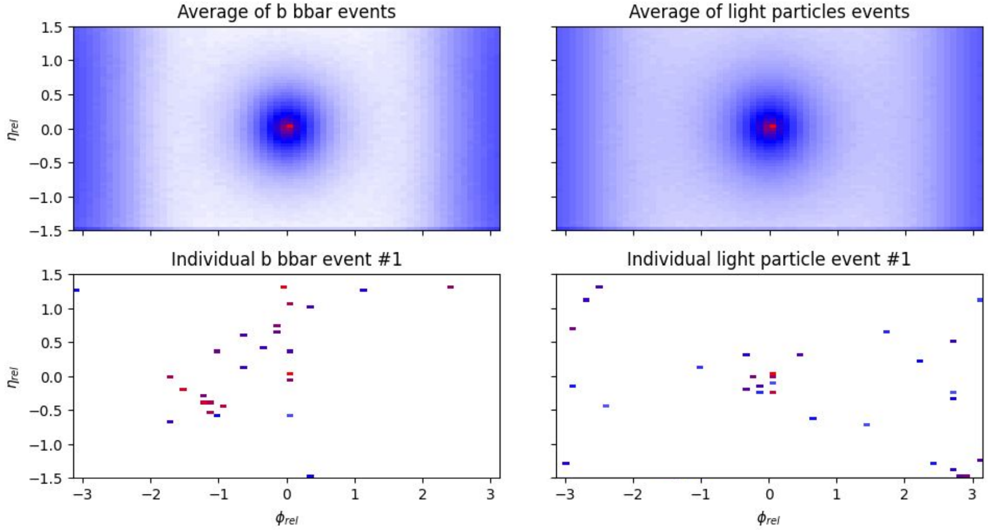


Image from [3]

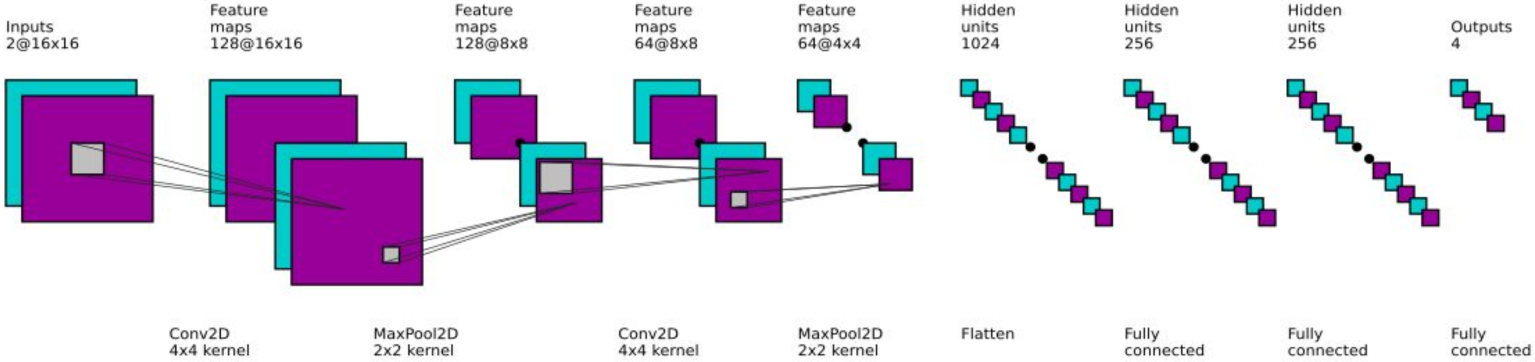


[3] H. Qu, C. Li, and S. Qian, "Particle Transformer for Jet Tagging", 2024.

The ability to tag events

Preliminary results

Main framework



Adapted from [4]'s top tagger

- Activation functions: tanh and softmax
- Epochs: 25
- Optimizer: Adam
- Learning rate: 0.00005, halving each 10 epochs
- Loss function: Categorical cross entropy

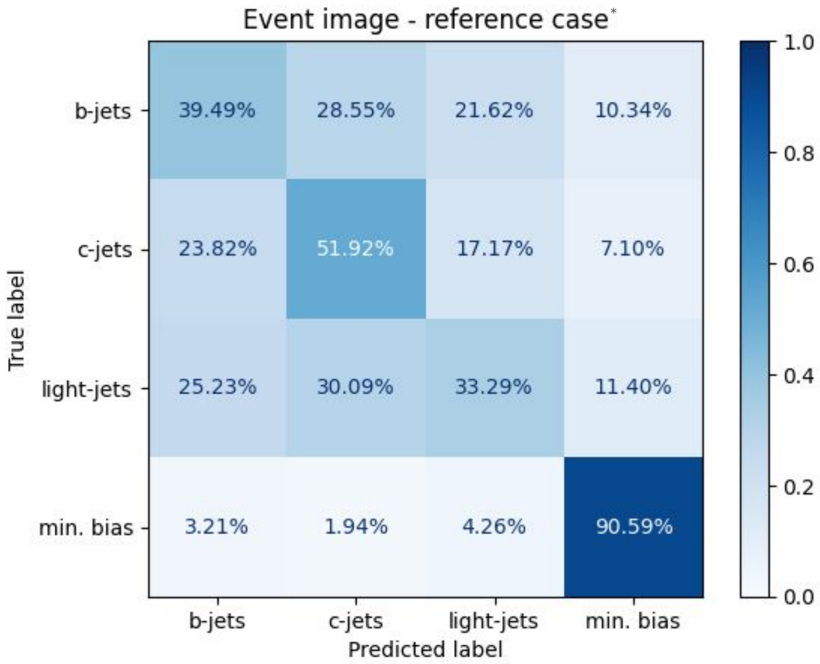
[4] G. Kasieczka et al., "The Machine Learning landscape of top taggers", SciPost Physics, vol. 7, no. 1, 2019.

Event images

*

*Confusion matrices normalized with respect to the true label

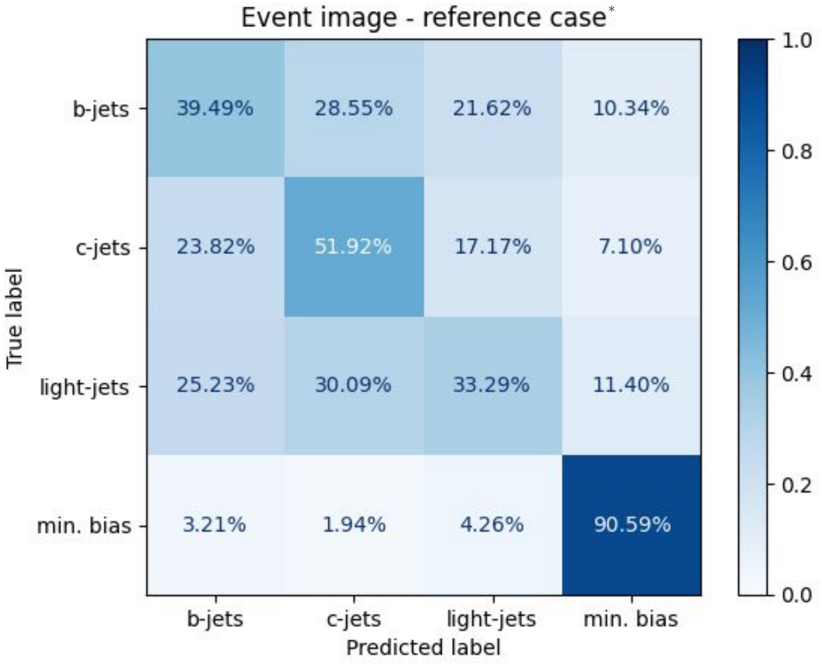
Event images



- 16x16 pixels
- R $\simeq$ 1 ($\Delta\eta_{rel} = \Delta\varphi_{rel} = 0.7$)
- Different $\hat{p}_T$ ranges mixed
- 2 color channels: p_T and charge

*Confusion matrices normalized with respect to the true label

Event images

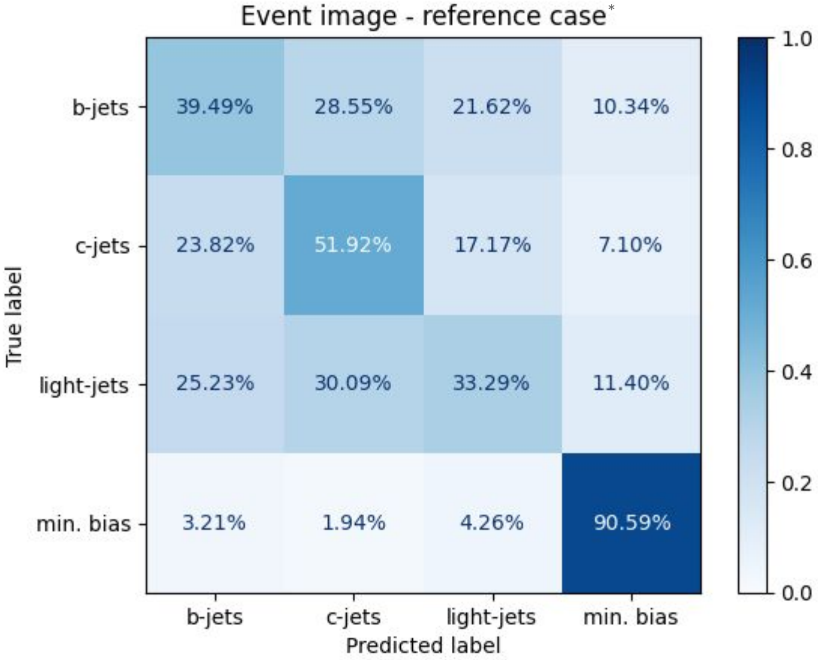


- 16x16 pixels
- $R \simeq 1$ ($\Delta\eta_{rel} = \Delta\varphi_{rel} = 0.7$)
- Different $\hat{p}_T$ ranges mixed
- 2 color channels: p_T and charge

➔ Flavor tagging needs further improvement

*Confusion matrices normalized with respect to the true label

Event images



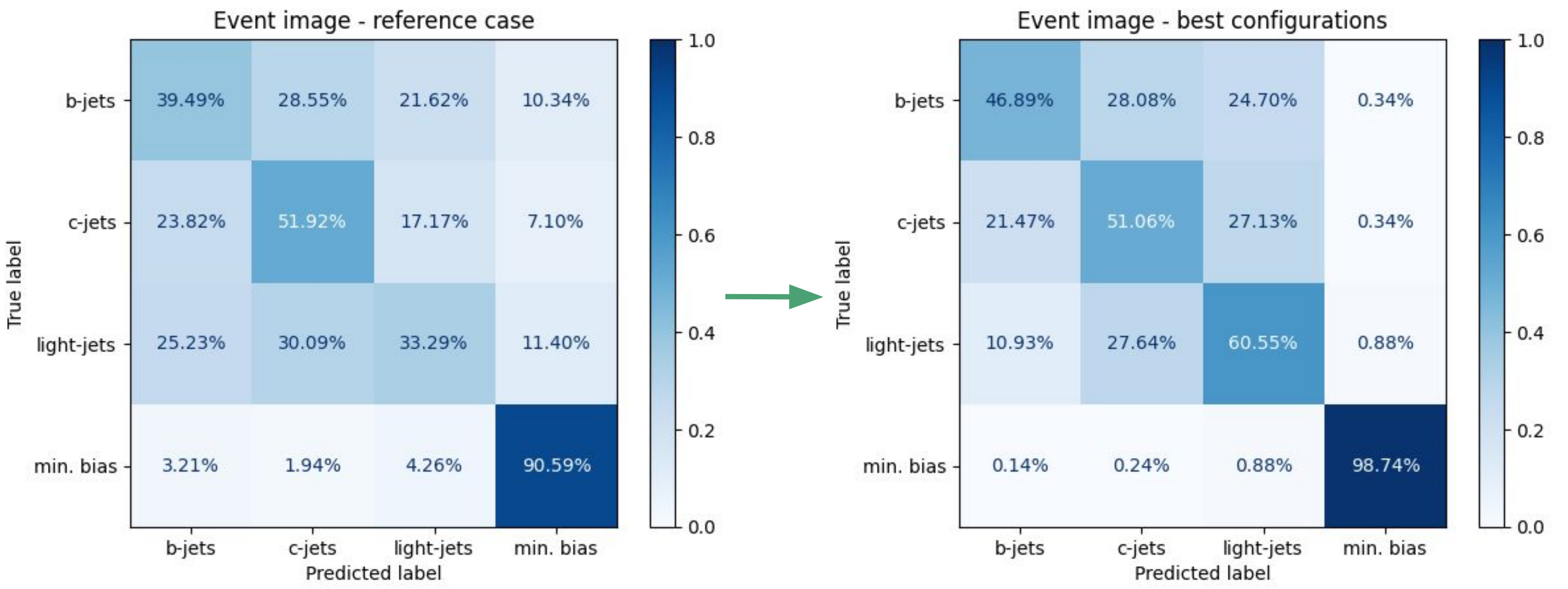
- 16x16 pixels
- R ≃ 1 ($\Delta\eta_{rel} = \Delta\varphi_{rel} = 0.7$)
- Different $\hat{p}_T$ ranges mixed
- 2 color channels: p_T and charge

➔ Flavor tagging needs further improvement

➔ Excels at background separation!

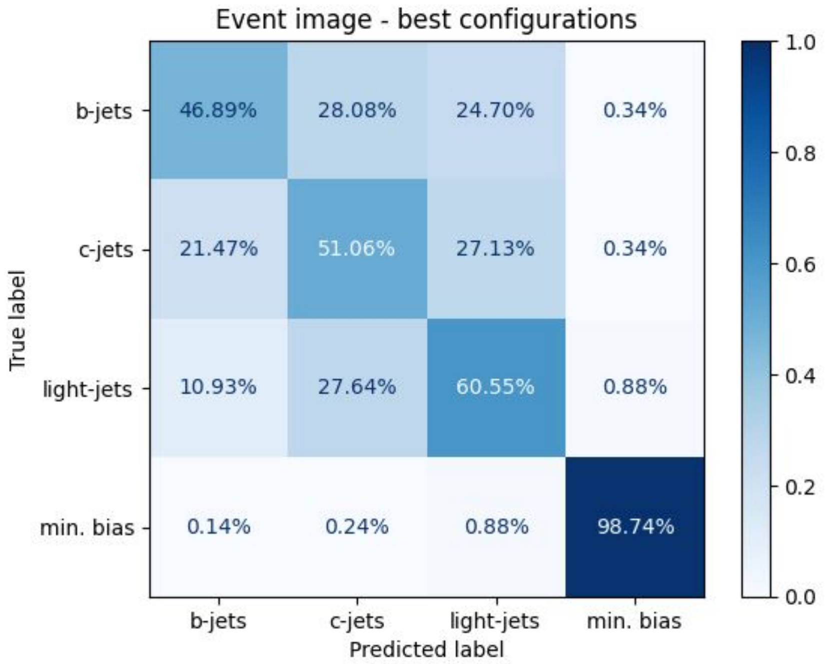
*Confusion matrices normalized with respect to the true label

Event images

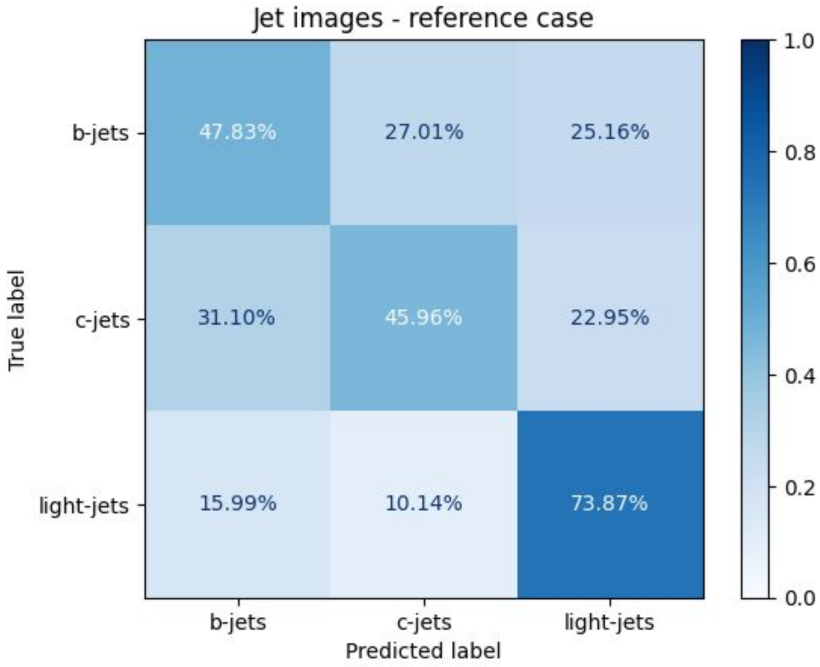


Event images

- $R \simeq 2.8$ ($\Delta\eta_{rel} = \Delta\varphi_{rel} = 2$)
- Highest $\hat{p}_T$ range (90 - 110 GeV/c)
- 5 color channels: p_T, charge, baryon, lepton and "meson" number

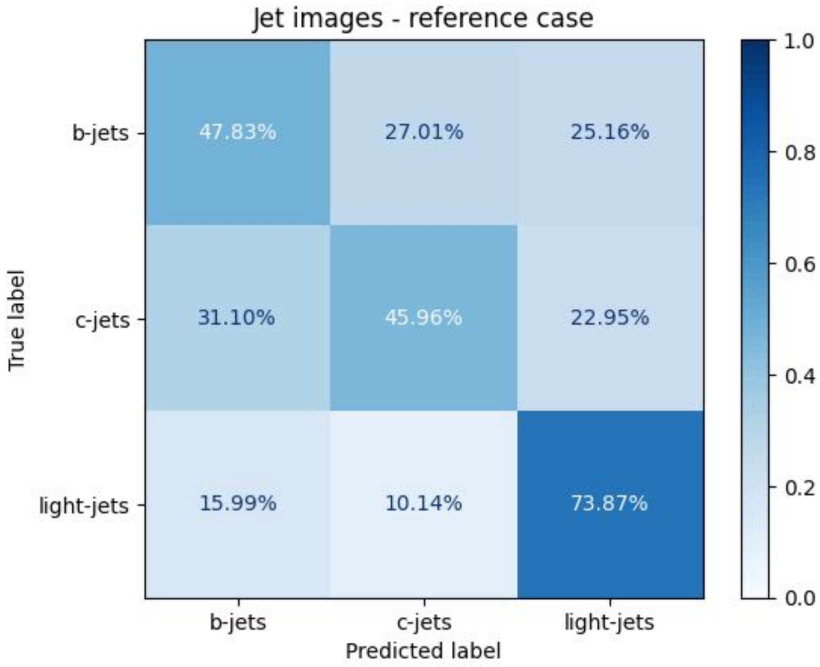


Jet images



- 16x16 pixels
- $R = \sqrt{2}$ ($\Delta\eta_{rel} = \Delta\varphi_{rel} = 1$)
- No cuts in p_T
- particle p_T and charge color channels

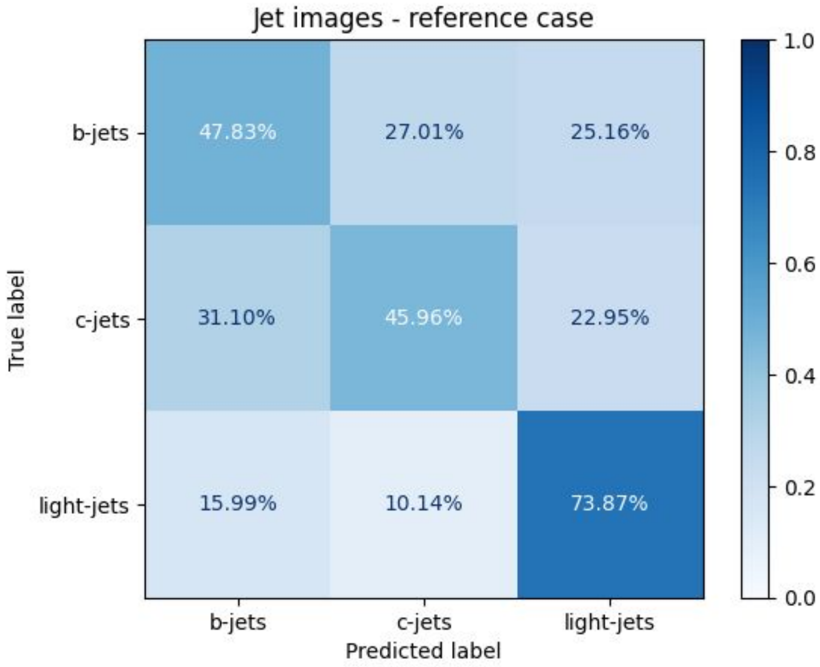
Jet images



- 16x16 pixels
- R = √2 ($\Delta\eta_{rel} = \Delta\varphi_{rel} = 1$)
- No cuts in p_T
- particle p_T and charge color channels

➔ Similar performance for the heavy jets

Jet images



- 16x16 pixels
- $R = \sqrt{2}$ ($\Delta\eta_{rel} = \Delta\varphi_{rel} = 1$)
- No cuts in p_T
- particle p_T and charge color channels

➔ Similar performance for the heavy jets

➔ Light tagging improved!

Insights from reconstructed jets

Discussion and outlook

Questions and (possible) answers

Questions and (possible) answers

- Shouldn’t the reconstructed jets yield a (even) better result?

Questions and (possible) answers

- Shouldn’t the reconstructed jets yield a (even) better result?
  - b and c jets are much more similar than light jets

Questions and (possible) answers

- Shouldn't the reconstructed jets yield a (even) better result?
  - b and c jets are much more similar than light jets
  - Results given by direct network outputs, discriminant(s) should improve[1]

Questions and (possible) answers

- Shouldn't the reconstructed jets yield a (even) better result?
  - b and c jets are much more similar than light jets
  - Results given by direct network outputs, discriminant(s) should improve[1]

$$D_{q_i} = log\left(\frac{p_{q_i}}{f_{q_j}p_{q_j}+(1-f_{q_j})p_{q_k}}\right)$$

Questions and (possible) answers

- Shouldn't the reconstructed jets yield a (even) better result?
  - b and c jets are much more similar than light jets
  - Results given by direct network outputs, discriminant(s) should improve[1]

$$D_{q_i} = log\left(\frac{p_{q_i}}{f_{q_j}p_{q_j}+(1-f_{q_j})p_{q_k}}\right)$$

  - Same low level variables as before, but jet properties could help

Questions and (possible) answers

- Shouldn't the reconstructed jets yield a (even) better result?
  - b and c jets are much more similar than light jets
  - Results given by direct network outputs, discriminant(s) should improve[1]

$$D_{q_i} = log\left(\frac{p_{q_i}}{f_{q_j}p_{q_j}+(1-f_{q_j})p_{q_k}}\right)$$

  - Same low level variables as before, but jet properties could help
- Where does this put event images?

Extracting jet information

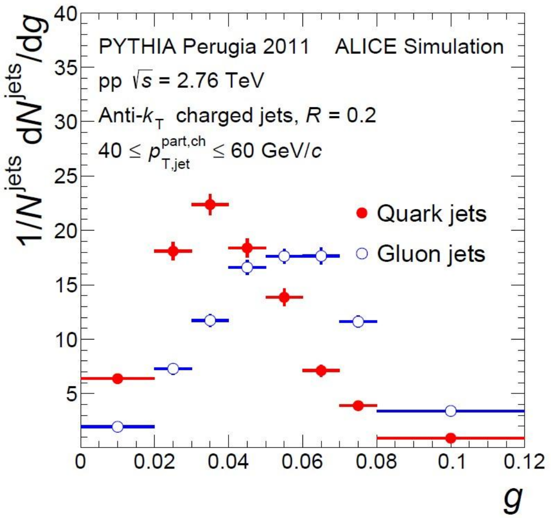
Extracting jet information

- Jet shape is sensitive to jet flavor[5]:

[5] The ALICE collaboration *et al.* "Medium modification of the shape of small-radius jets in central Pb-Pb collisions at √2.76 TeV". *J. High Energ. Phys.* **2018**, 139 (2018).

Extracting jet information

- Jet shape is sensitive to jet flavor[5]:

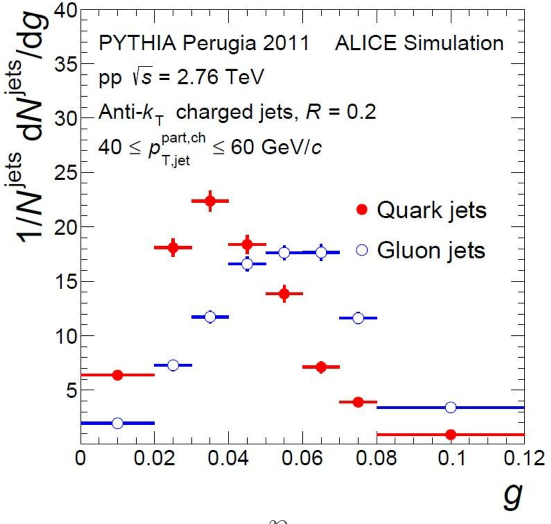


$$g = \sum_{i \in \mathrm{jet}} \frac{p_{T,i}}{p_{T,\mathrm{jet}}} \Delta R_{\mathrm{jet},i}$$

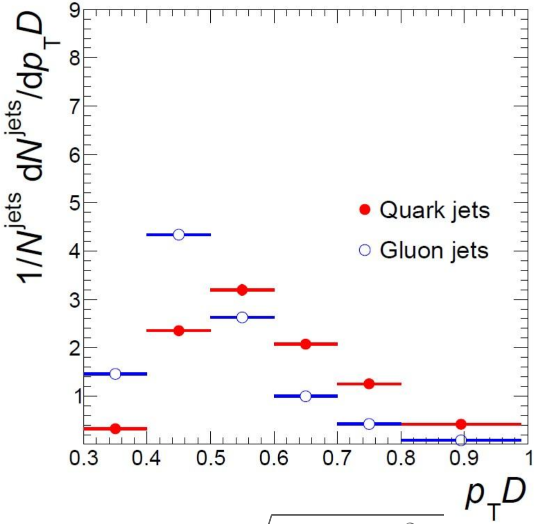
[5] The ALICE collaboration *et al.* "Medium modification of the shape of small-radius jets in central Pb-Pb collisions at √2.76 TeV". *J. High Energ. Phys.* **2018**, 139 (2018).

Extracting jet information

- Jet shape is sensitive to jet flavor[5]:



$$g = \sum_{i \in \mathrm{jet}} \frac{p_{T,i}}{p_{T,\mathrm{jet}}} \Delta R_{\mathrm{jet},i}$$



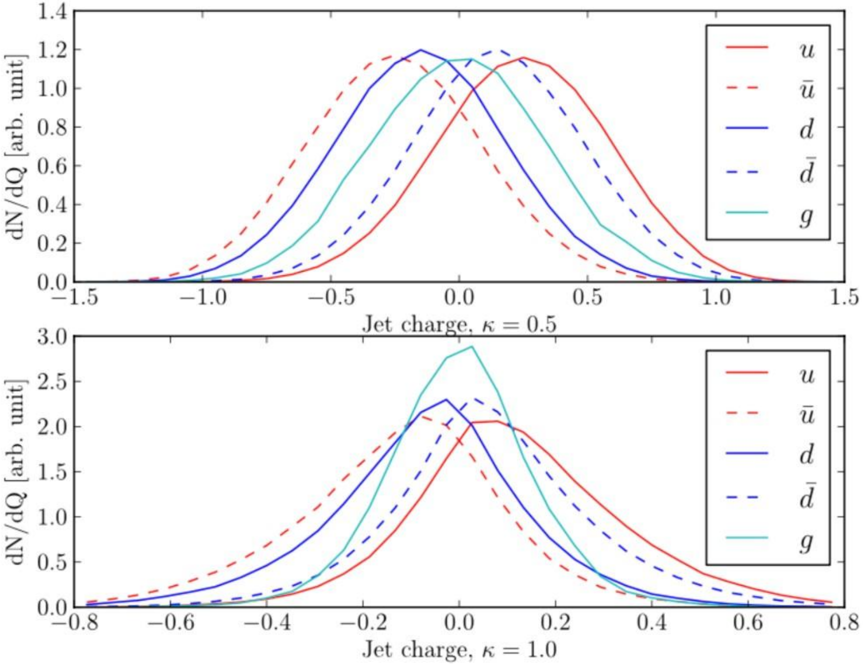
$$p_T D = \frac{\sqrt{\sum_{i \in \mathrm{jet}} p_{T,i}^2}}{\sum_{i \in \mathrm{jet}} p_{T,i}}$$

[5] The ALICE collaboration *et al.* "Medium modification of the shape of small-radius jets in central Pb-Pb collisions at √2.76 TeV". *J. High Energ. Phys.* **2018**, 139 (2018).

Extracting other jet information

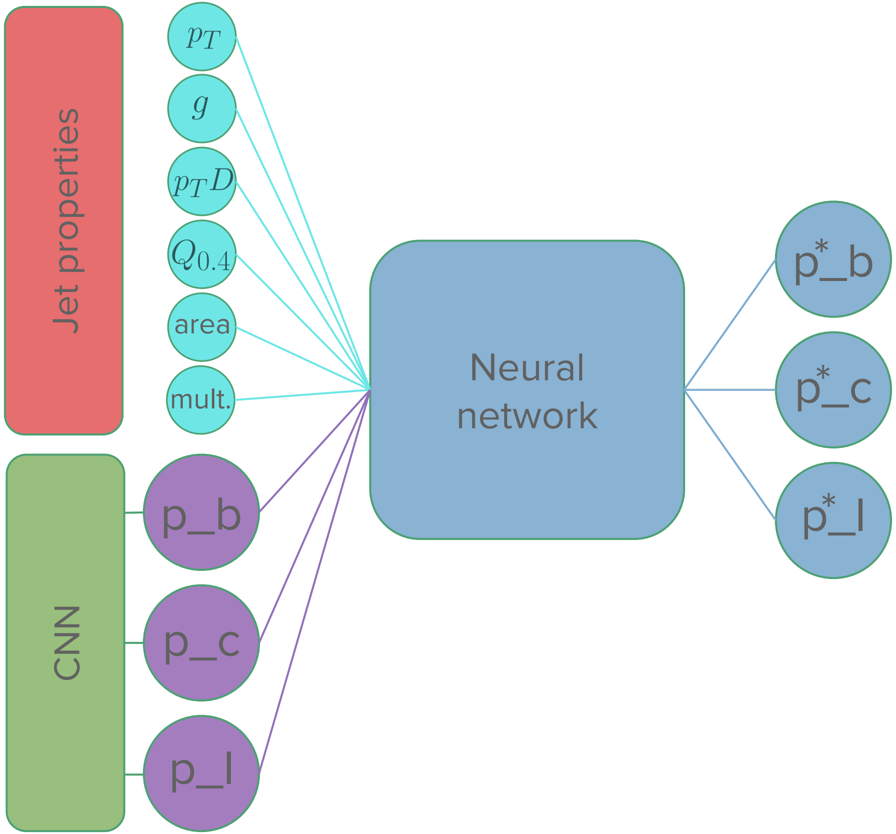
- Other jet properties as well[6]

$$Q_k = \sum_{i \in \text{jet}} \left(\frac{p_{T,i}}{p_{T,\text{jet}}} \right)^k Q_i$$

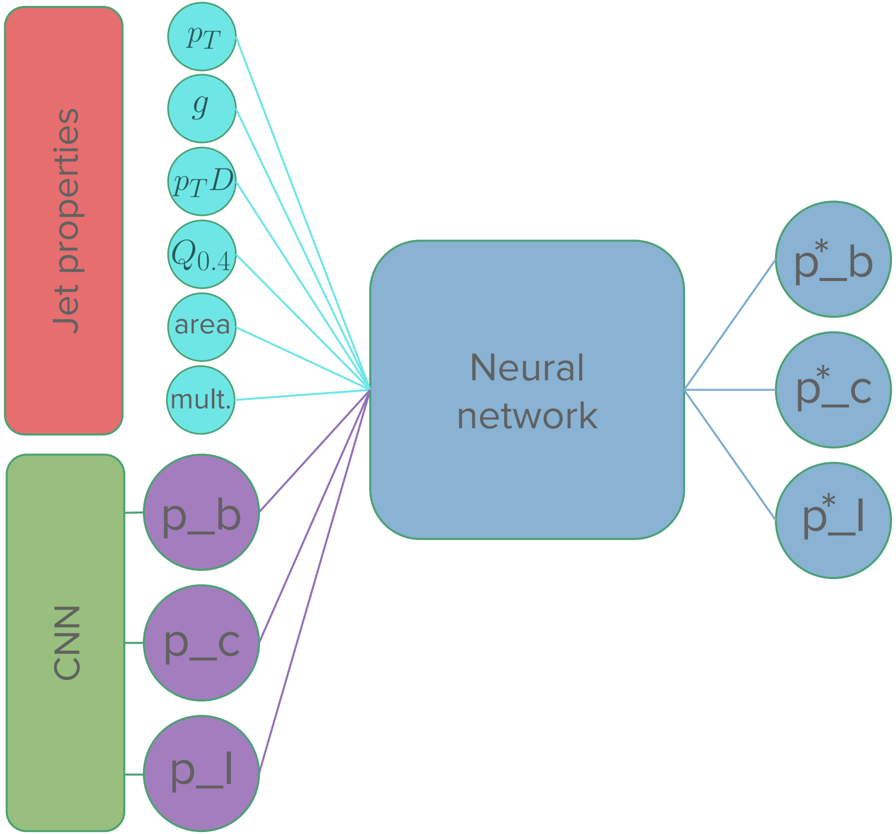


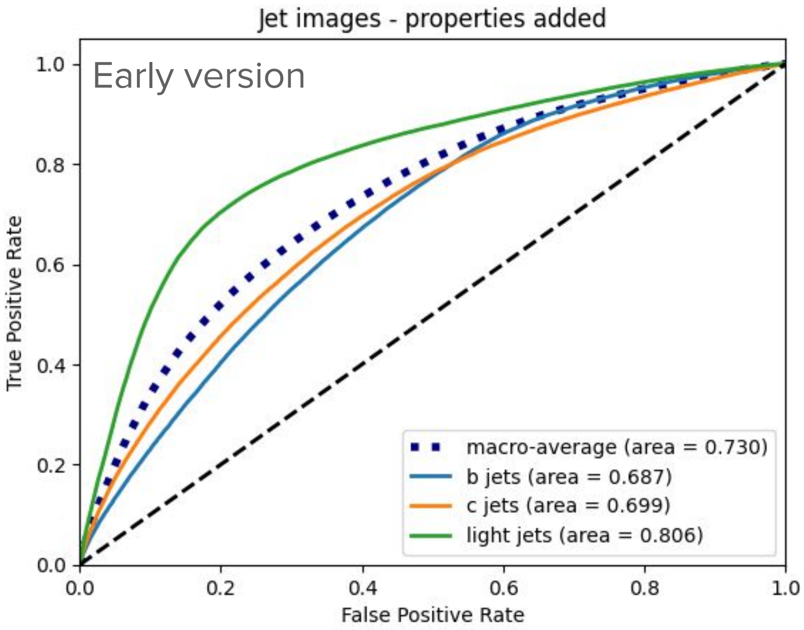
[6] D. Krohn, M. D. Schwartz, T. Lin, and W. J. Waalewijn, "Jet charge at the LHC", Phys. Rev. Lett., vol. 110, p. 212001, May 2013.

Adding a second network



Adding a second network





Outlook

Outlook

- Improve the secondary network

Outlook

- Improve the secondary network
- Adapt observables to event landscape

Outlook

- Improve the secondary network
- Adapt observables to event landscape
- Hope for the best!

Thank you!

Finance Code 88887.674361/2022-00,
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil

Grant #2022/02989-9,
Grant #2025/28345-9,
São Paulo Research Foundation (FAPESP)

The opinions, hypotheses and conclusions or recommendations expressed in this material are the responsibility of the author(s) and do not necessarily reflect the views of FAPESP and CAPES.

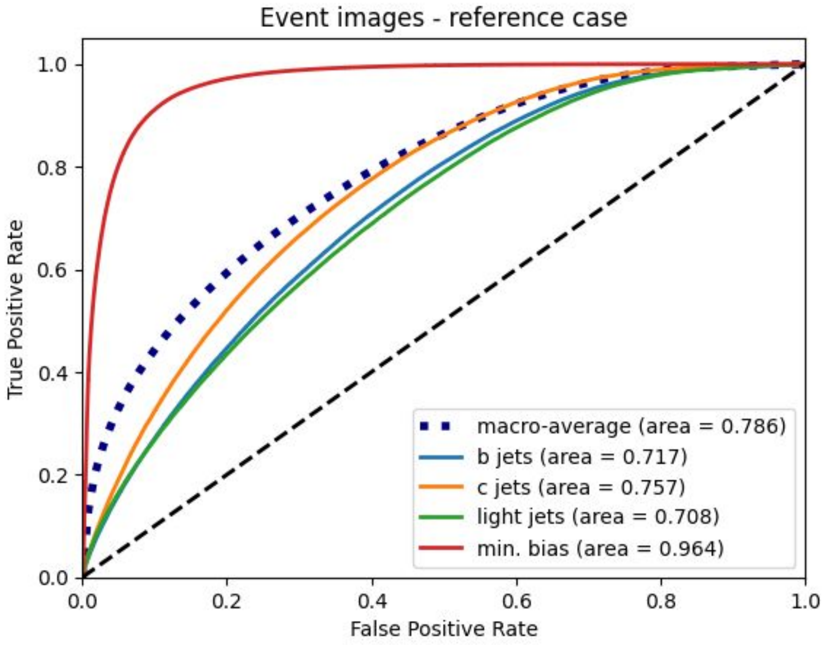
Backup

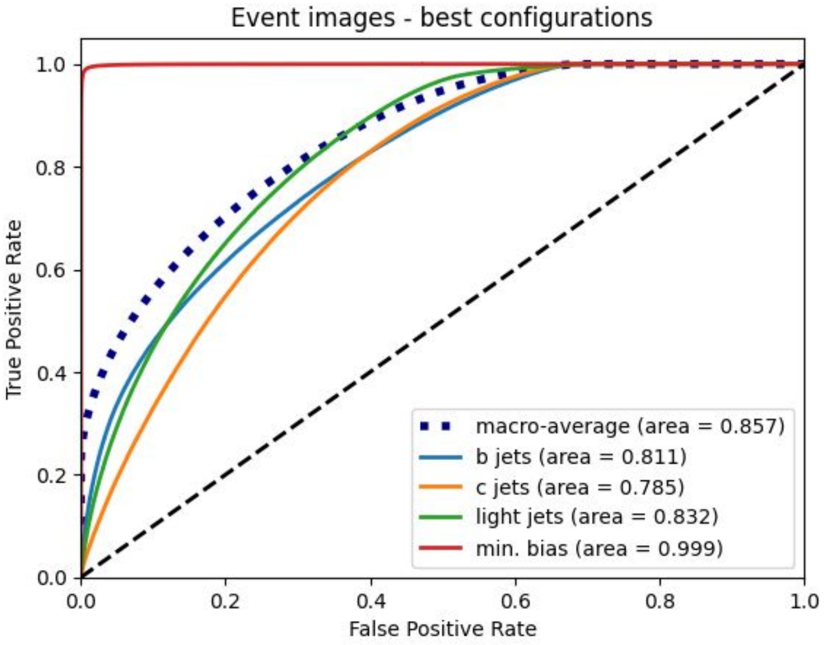
The database

- Simulated collisions
  - Pythia 8.3 Monash Tune with CT10 PDFs
  - p+p with 13 TeV
- Selected processes
  - One process per event
  - Production of bottom, charm, light particles (light quark + gluons)
  - Minimum bias for background
- Defining our jets
  - Charged final state particles (detectable)
  - 50 most energetic particles
  - Hard processes (except background)
- Cuts in $\hat{p}_T$ phase space
  - 20 to 30 GeV
  - 40 to 60 GeV
  - 90 to 110 GeV

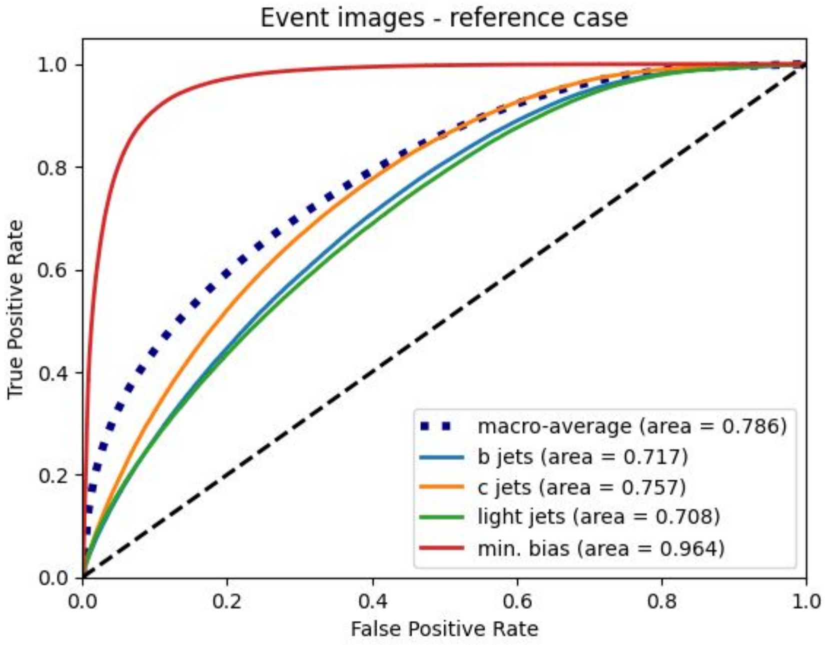
ROC curves

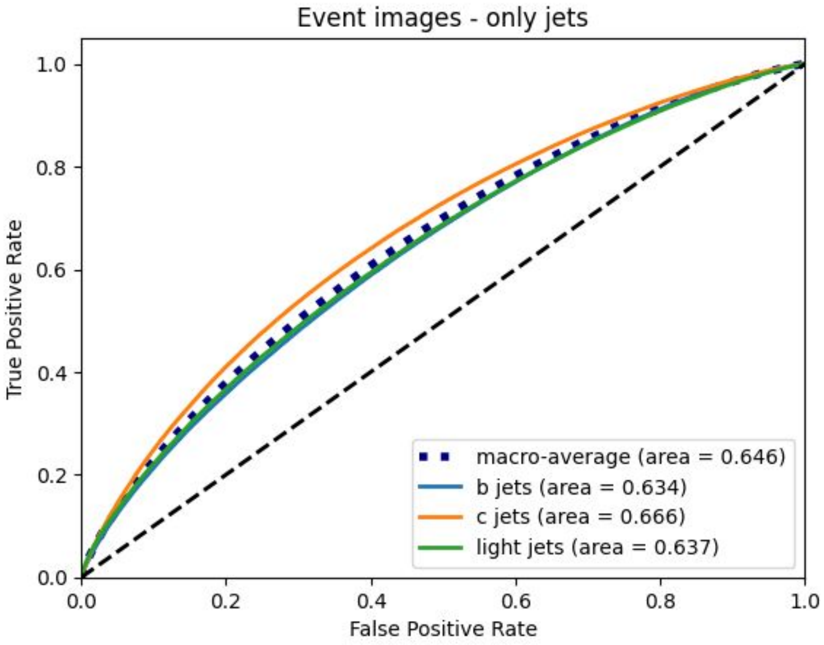
Event images



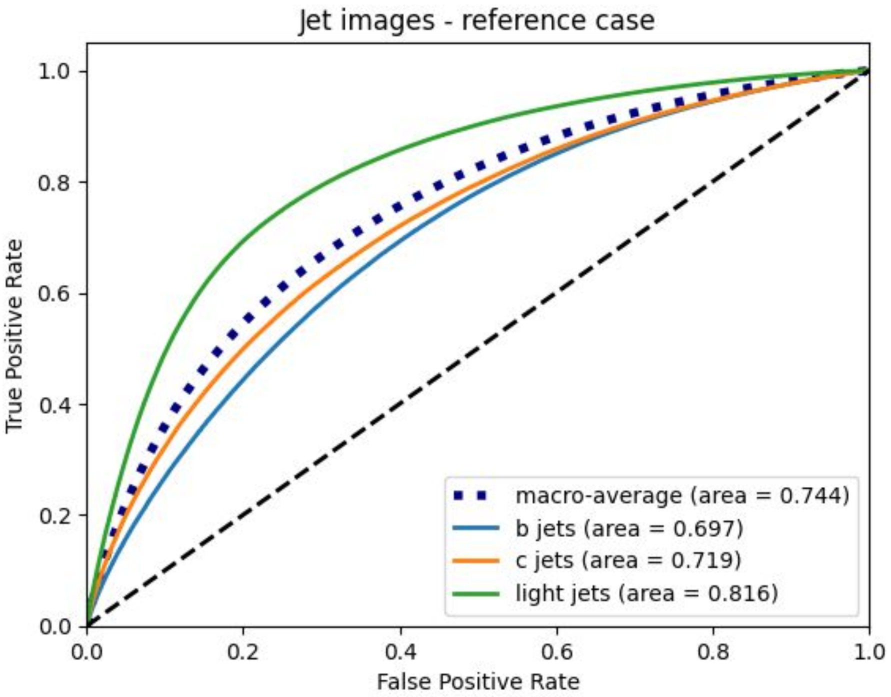


Event images





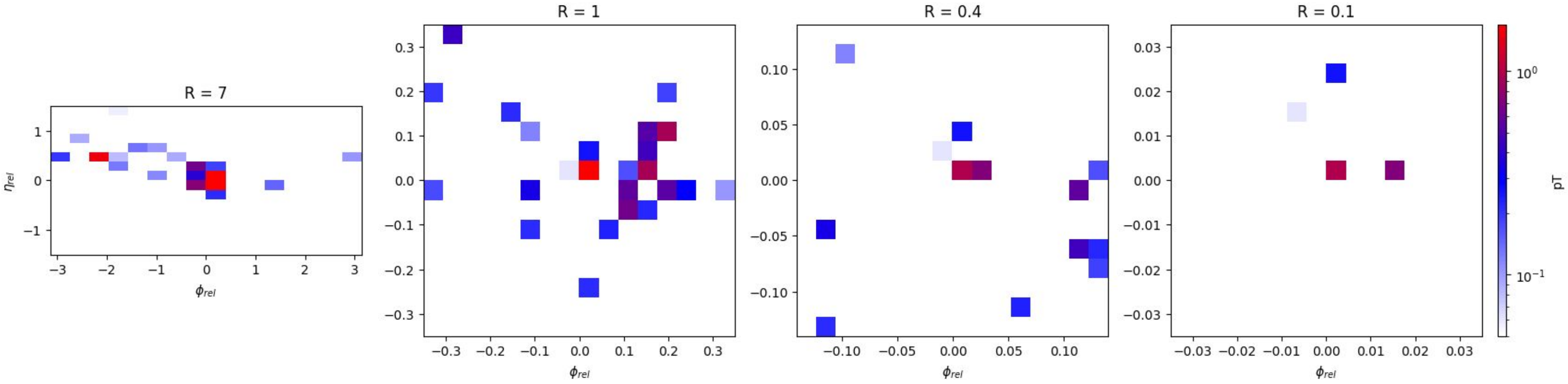
Jet images


Jet images - reference case
1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate
macro-average (area = 0.744)
b jets (area = 0.697)
c jets (area = 0.719)
light jets (area = 0.816)

Studies in event images configuration

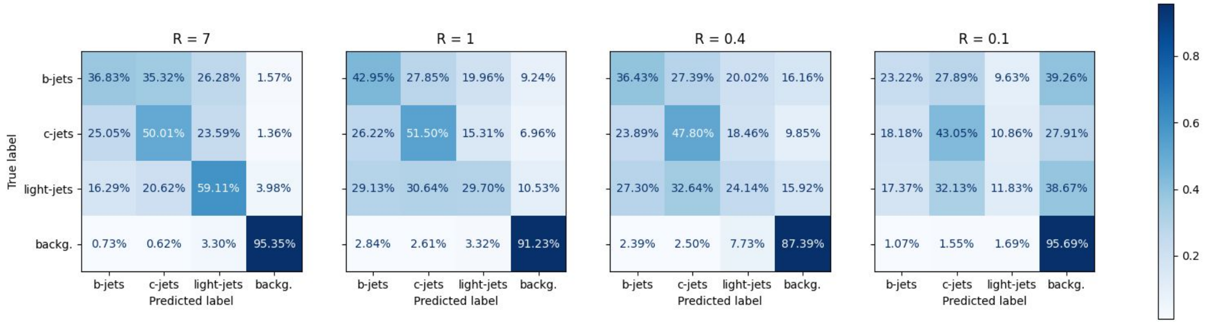
Applied models

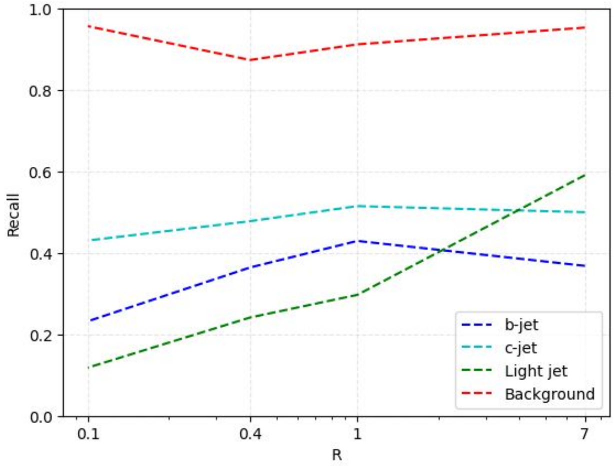
➔ Configuration studies
  - Images zoom (R factor)



Applied models

➔ Configuration studies
  - Images zoom (R factor)



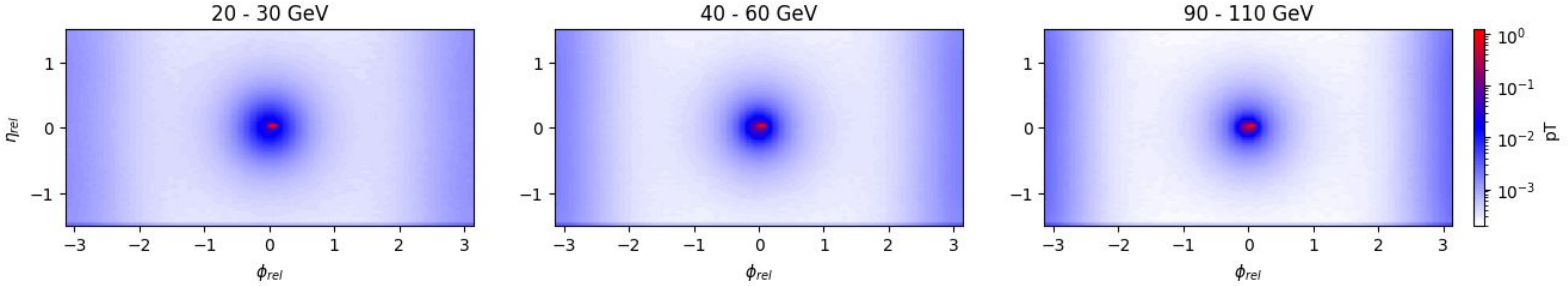


Zoom →

← Performance

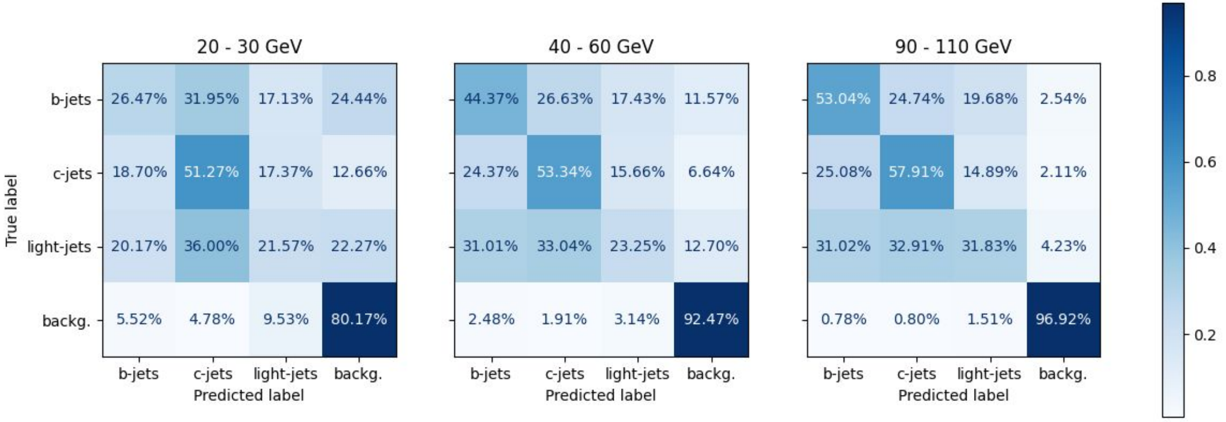
Applied models

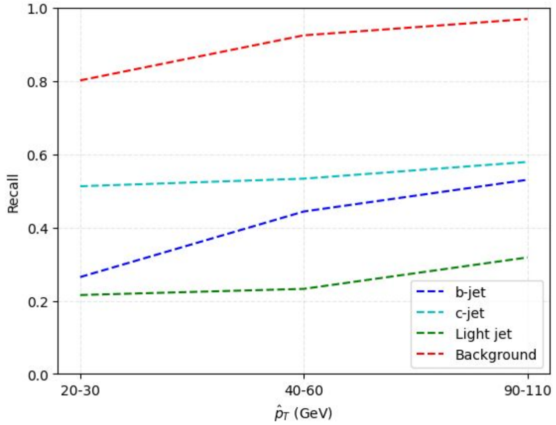
- ➔ Configuration studies
  - ◆ Separation by $\hat{p}_T$



Applied models

- ➔ Configuration studies
  - ◆ Separation by $\hat{p}_T$

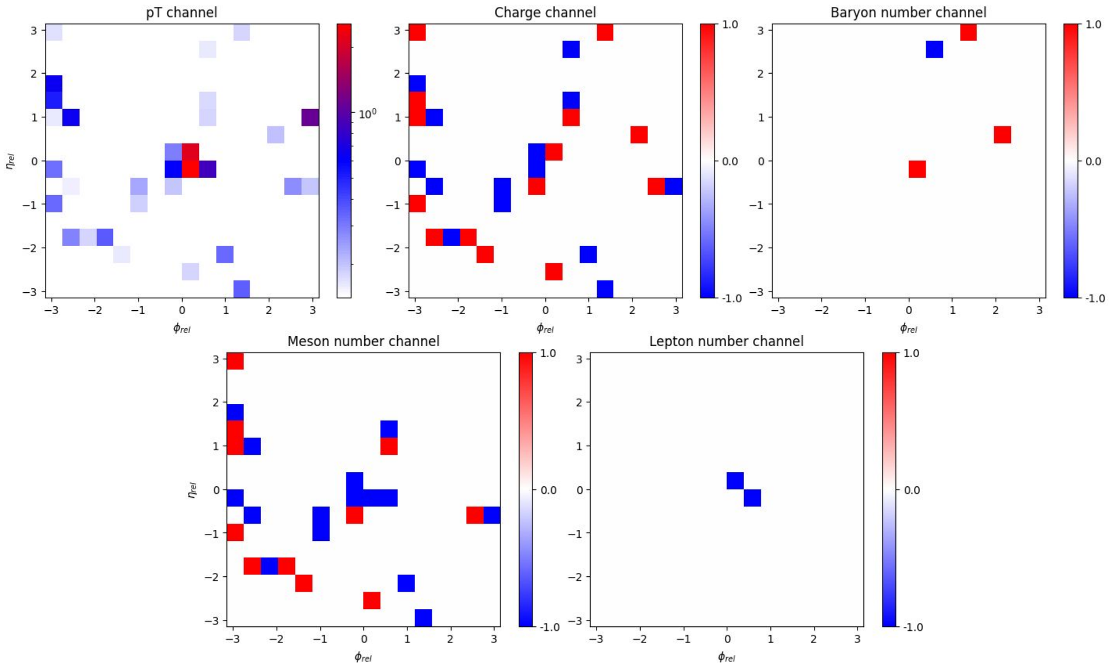




Higher energy ➔ More particles ➔ Better results

Applied models

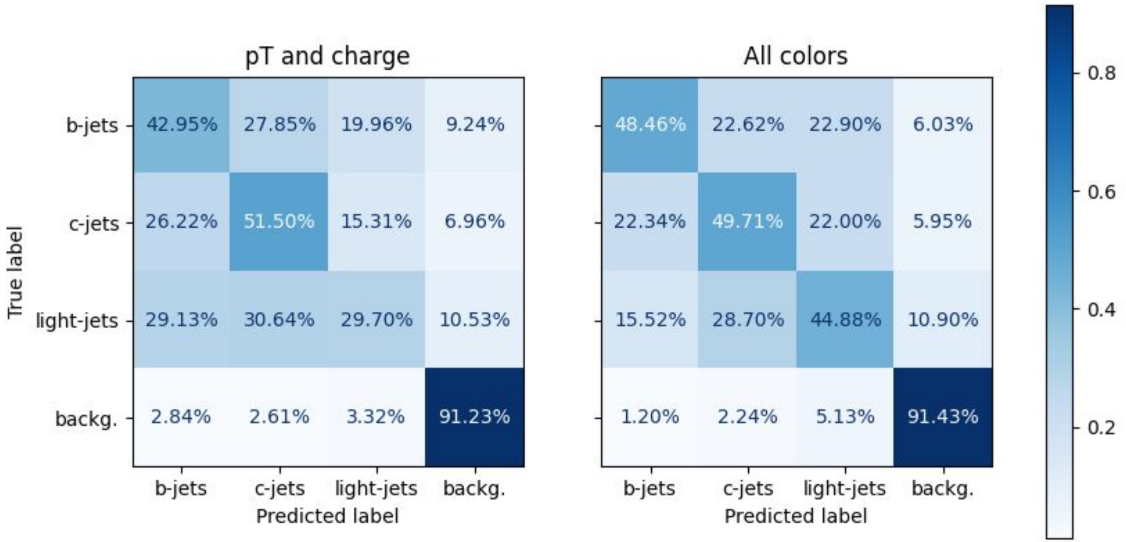
- Configuration studies
  - Addition of colors (particle identification)



Applied models

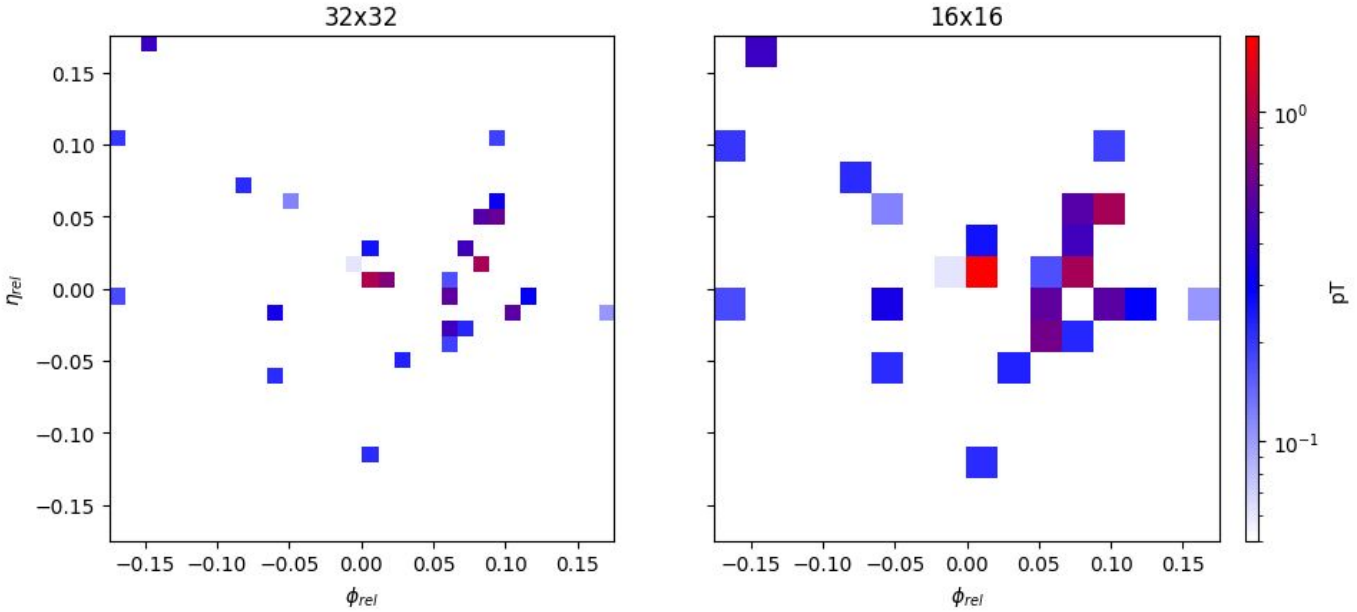
➔ Configuration studies
  ◆ Addition of colors (particle identification)

- Light jets reach the same level as heavy jets
- The recall on b-tagging gets better
- False positive rates of several classes improve as well



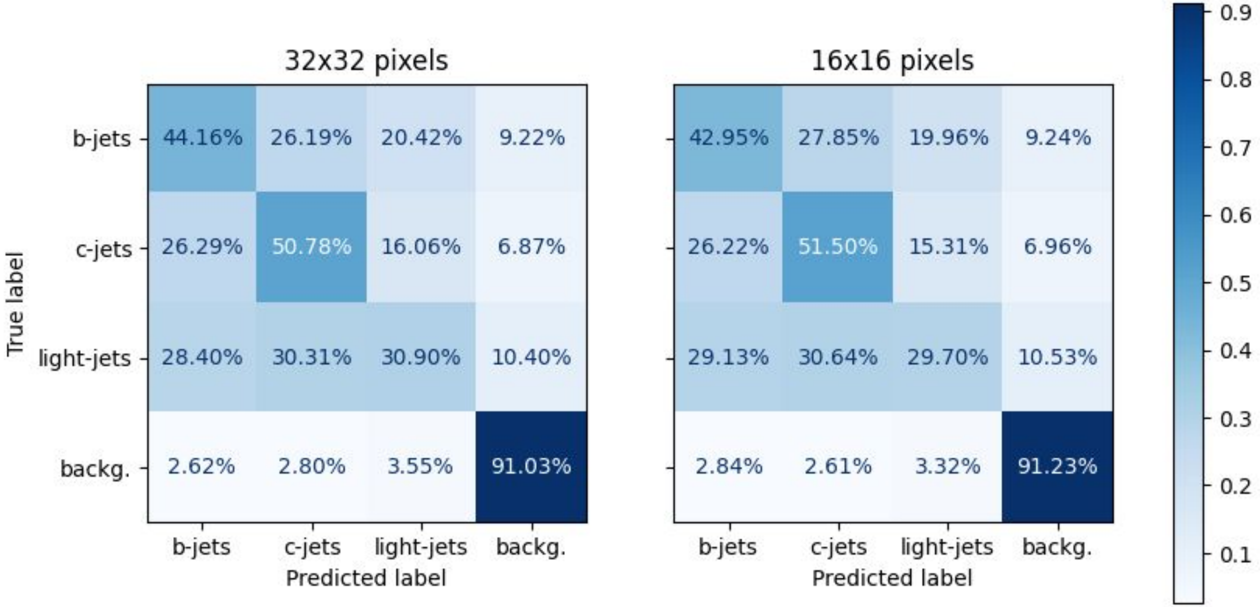
Applied models

- ➔ Image settings
  - ◆ Pixel resolution



Applied models

- ➔ Image settings
  - ◆ Pixel resolution



- ➔ Higher granularity improves the performance too little for the computational cost