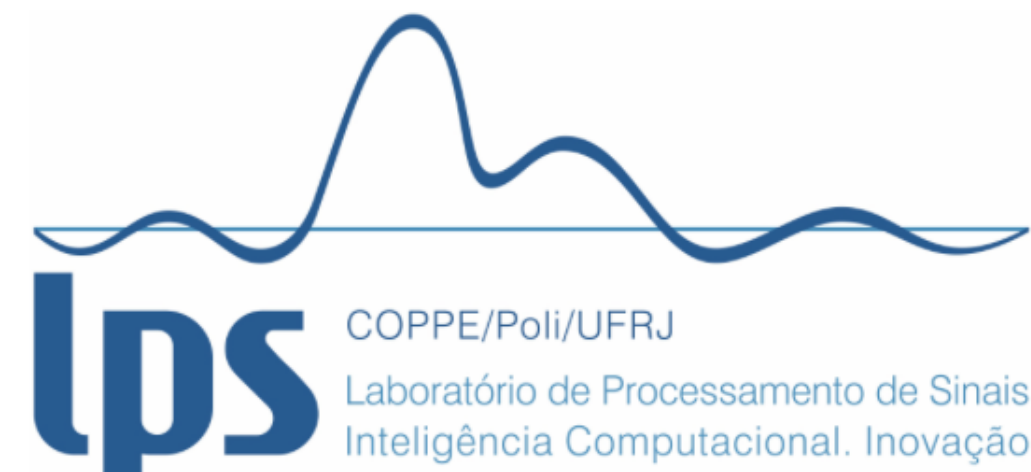


Leitura Multi-Anodal

Usando um Fotomultiplicadora Multi-Anodo e Aprendizagem Profunda para Tornar Mais Fina a Granularidade do Calorímetro Hadrônico TileCal do ATLAS

P. Gaspar, J. M. Seixas
12.07.2021



Projeto: Física Experimental de Altas Energias e Tecnologias Associadas

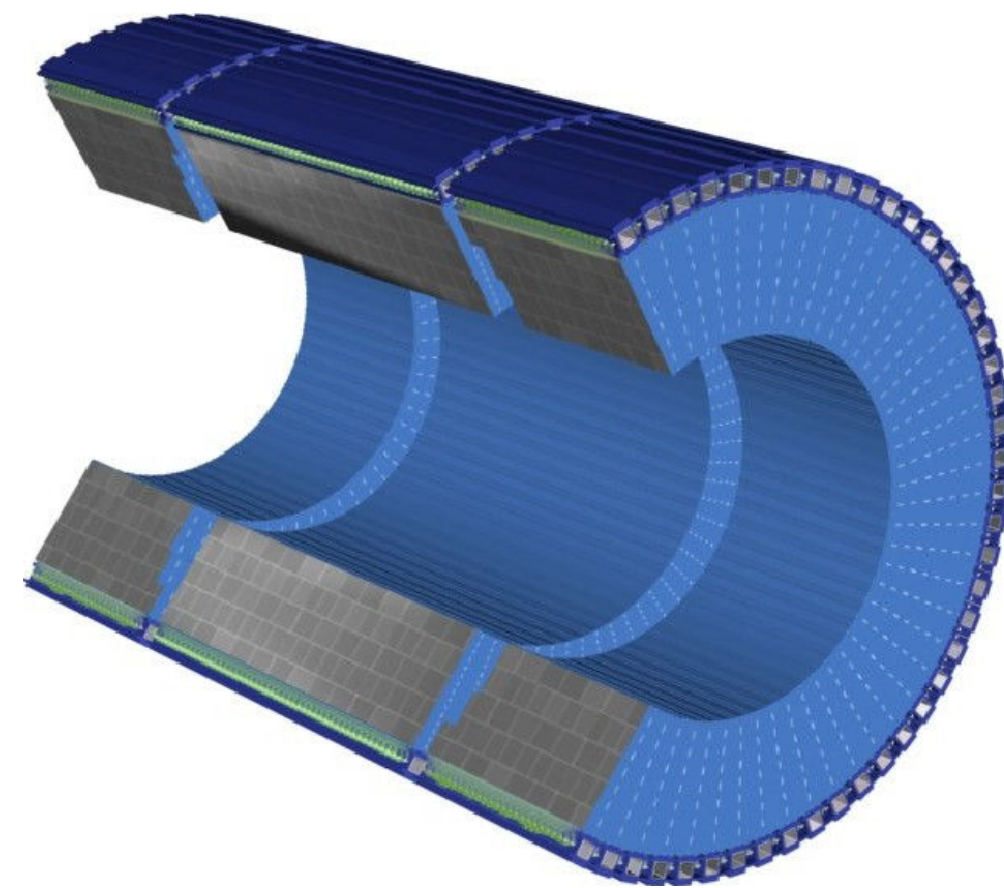
- Projeto em andamento em conjunto com o TileCal (LPS - COPPE/Poli/UFRJ & CERN).
- Combinação de processamento de sinais e aprendizado de máquina em experimentos de HEP.
- Spin-offs:
 - Possibilidade de aplicação da tecnologia em outros detectores, atuais e novos.
 - Metodologia similar pode ser aplicada em outras frentes; algumas aplicações já estão sendo feitas em medicina (TB, Covid-19).



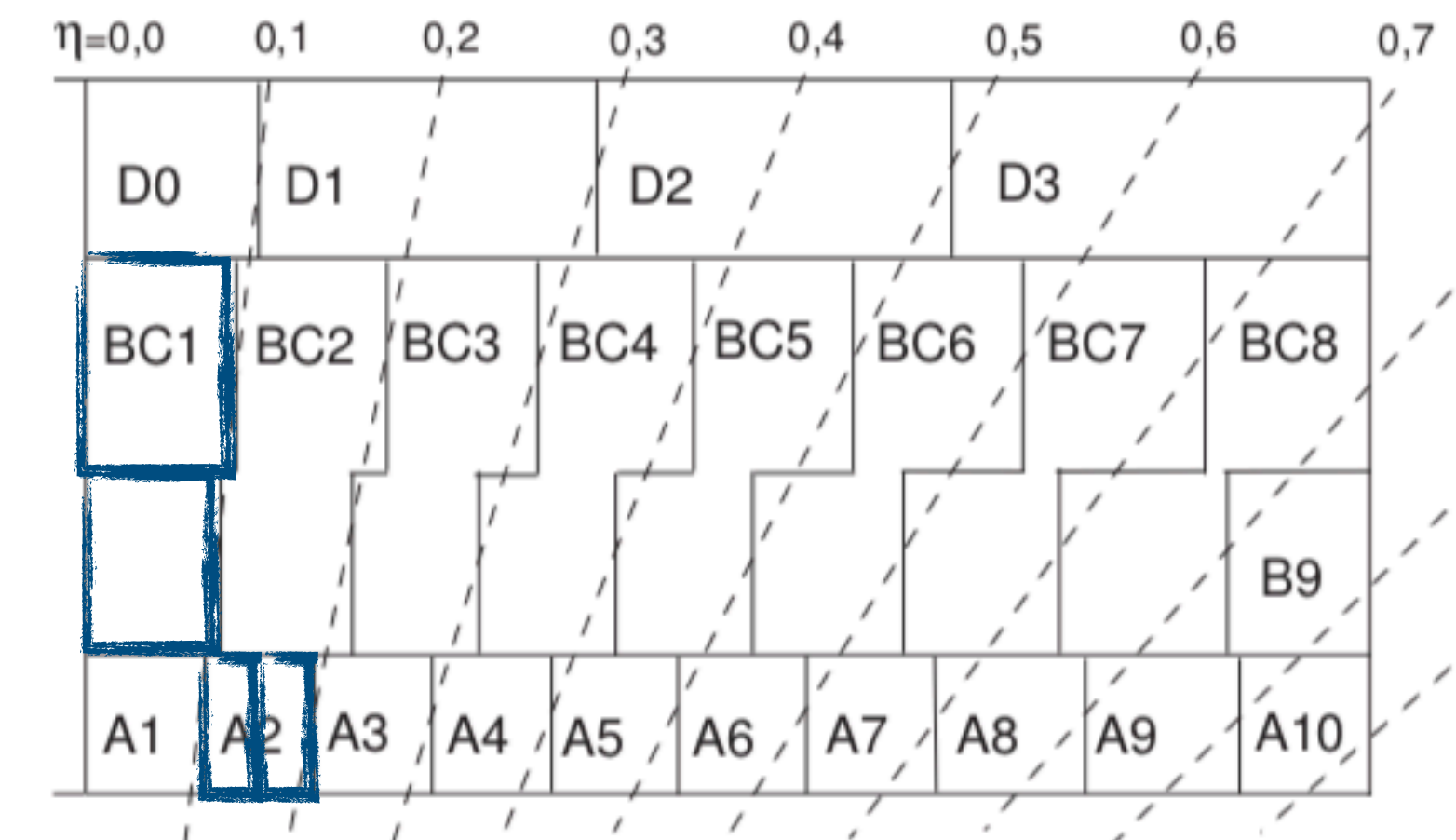
Objetivo

- Tornar mais fina a granularidade do TileCal através do uso de uma fotomultiplicadora de leitura Multi Anodal (MA-PMT).
- Associar as imagens formadas na grid de pixels da MA-PMT à regiões específicas dentro das células do calorímetro.
- Tornar a granularidade atual das camadas A e BC duas vezes mais fina, **sem modificar a estrutura mecânica do TileCal**.

Módulos do TileCal



Segmentação dos Módulos

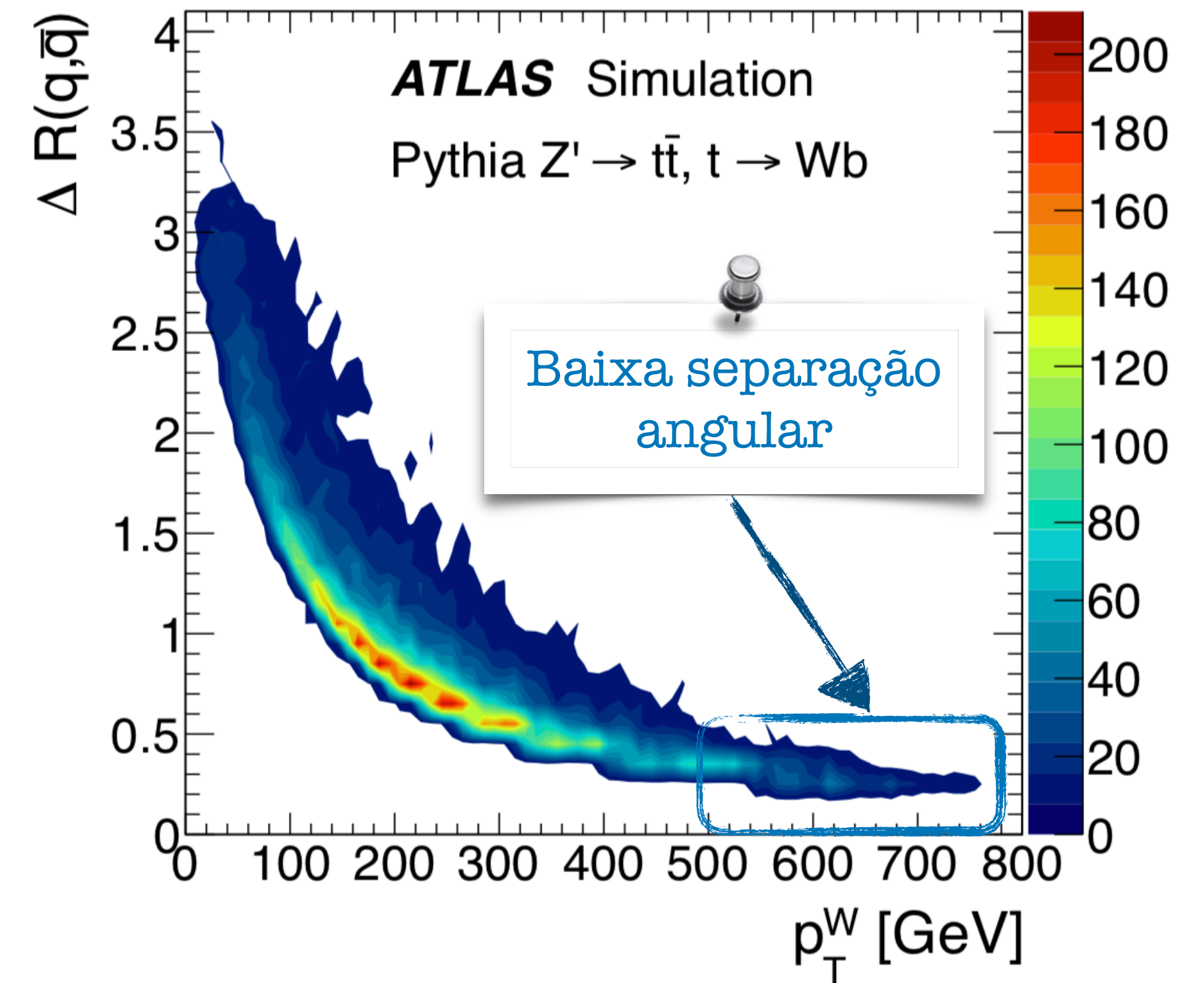
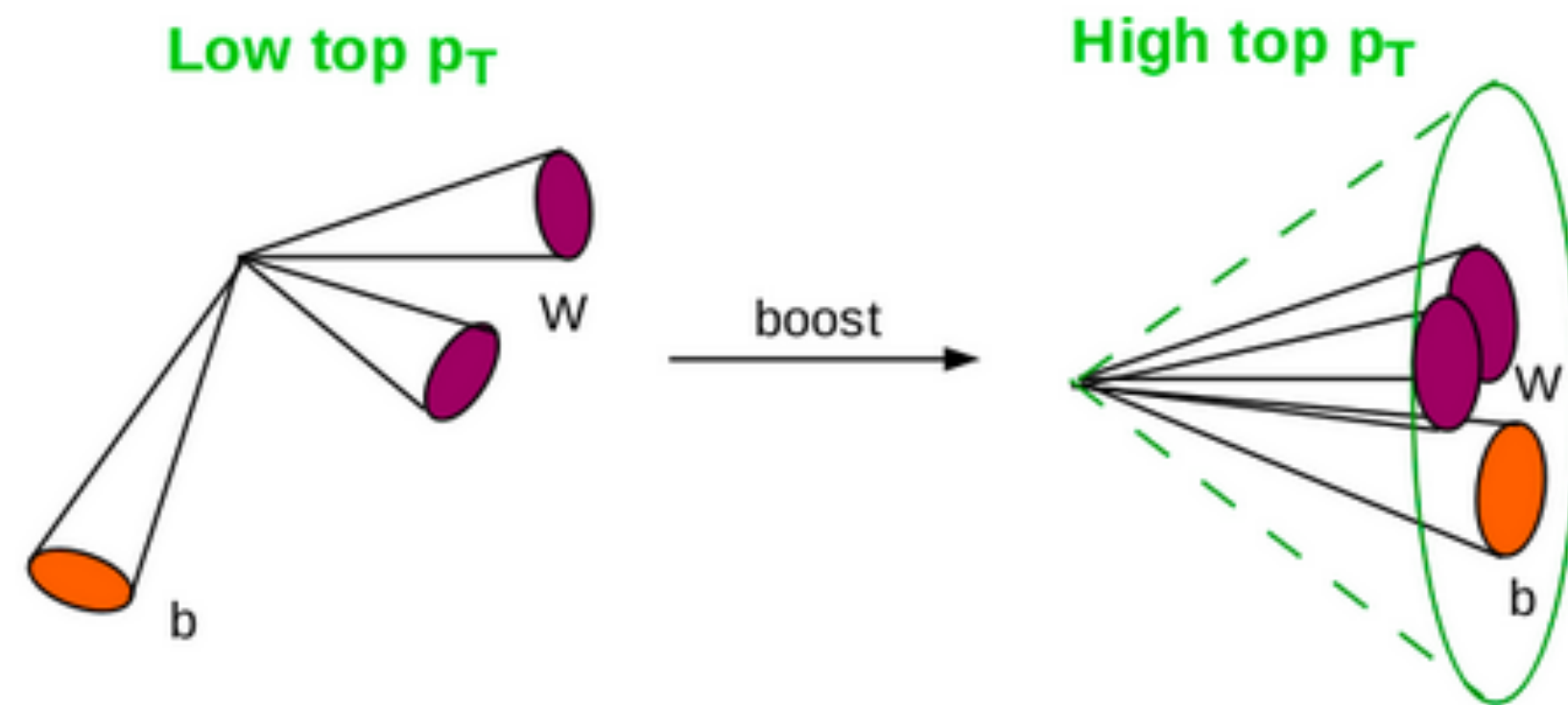


- O sinal temporal coletado pela PMT de anodo único, permite separar os dados, das células A e BC, em duas novas sub-regiões.

Essa separação permite criarmos um conjunto de dados anotados e assim olhar para o problema sobre a perspectiva do **aprendizado supervisionado**.

Motivação

- *Boosted Jets* no ambiente de alta luminosidade do LHC (HL-LHC).
- Quarks produzidos com um *boost* em seu momento.
- Seus produtos de decaimento possuem momento na mesma direção.



- Ângulo de separação entre dois quarks.
- Cada vez menor com o aumento de energia.

Introdução

- Cada célula do calorímetro é lida por um grupo de fibras responsáveis por transportar o sinal luminoso até o tubo fotomultiplicador localizado no final do módulo.
- Na configuração atual, cada célula é lida por dois tubos fotomultiplicadores.
- Para o a nova cadeia de leitura, substituímos uma das fotomultiplicadoras de anodo único por uma MA-PMT

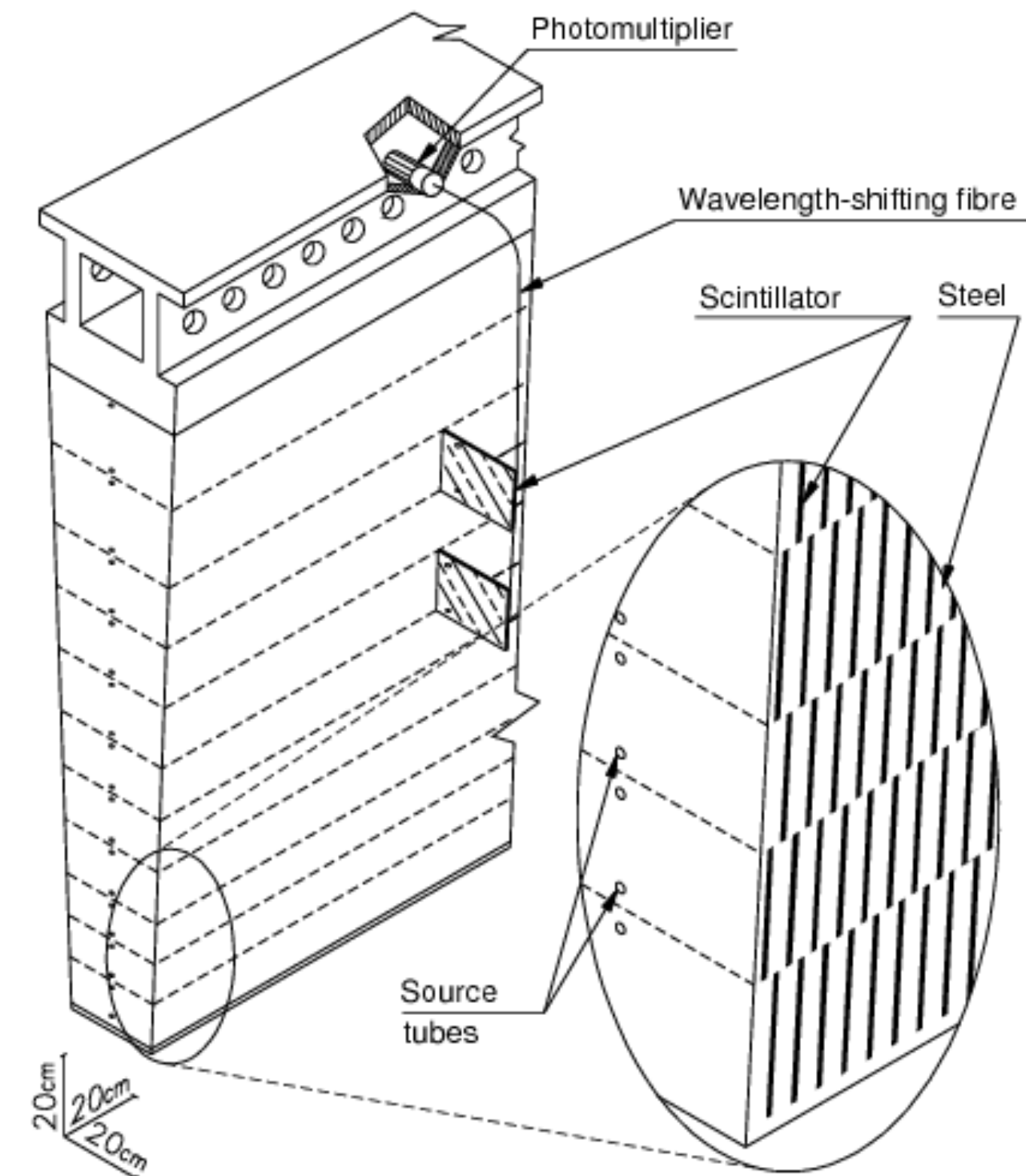
MA-PMT



Agrupamento de Fibras



Módulo do TileCal

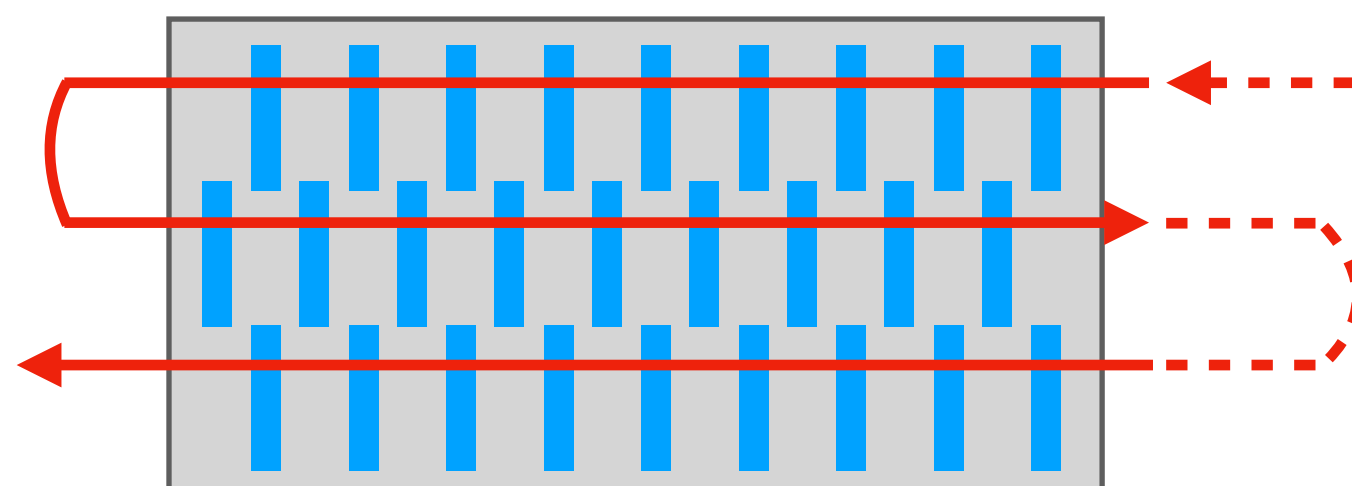


No novo arranjo experimental, o mesmo sinal é lido por dois tubos fotomultiplicadores com características diferentes.

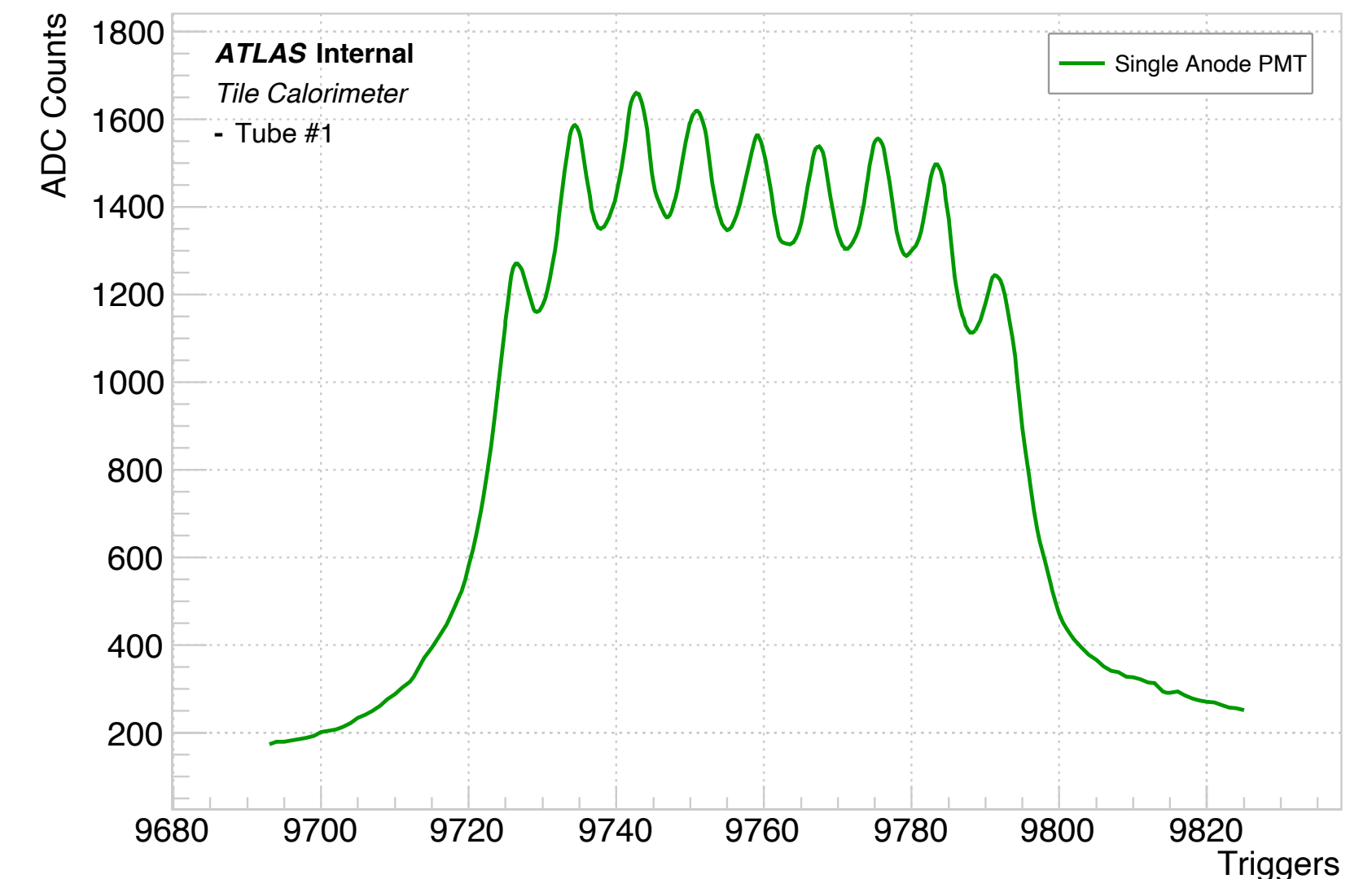
Conjunto de Dados

- Durante o processo de calibração, uma fonte radioativa de Cs137 passa por todos os módulos a fim de checar a qualidade e uniformidade da resposta de cada uma das telhas do calorímetro.
- Usamos o sinal temporal coletado pela PMT de anodo único, juntamente com o conhecimento da geometria do detector, para criar o conjunto de dados de treinamento.
- **Exemplo:**
 - Usar o perfil de sinal temporal para separar a célula A12 em duas novas sub-células (esquerda e direita).

Cesium Scan Path in A12 Cell

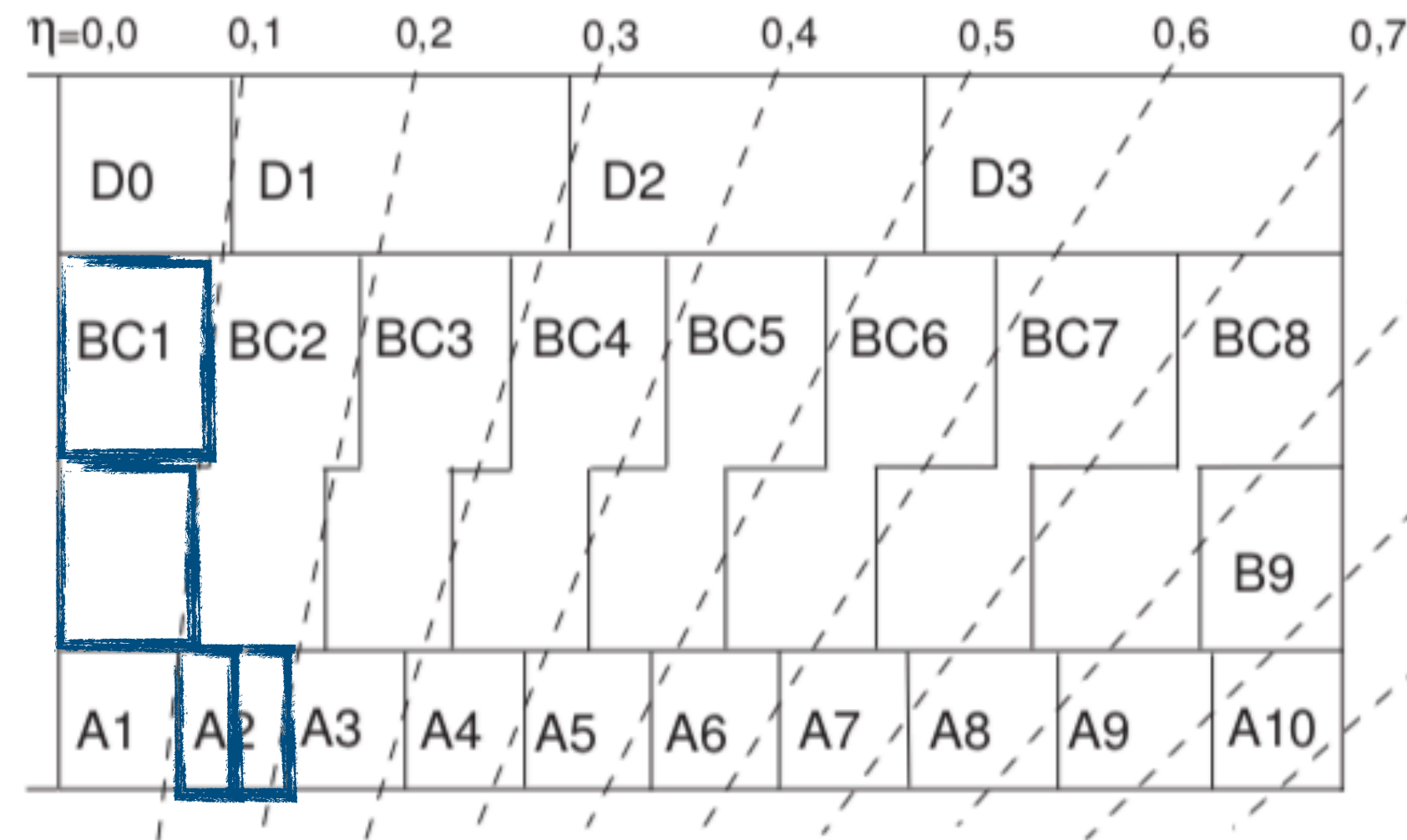


Perfil de Sinal Temporal

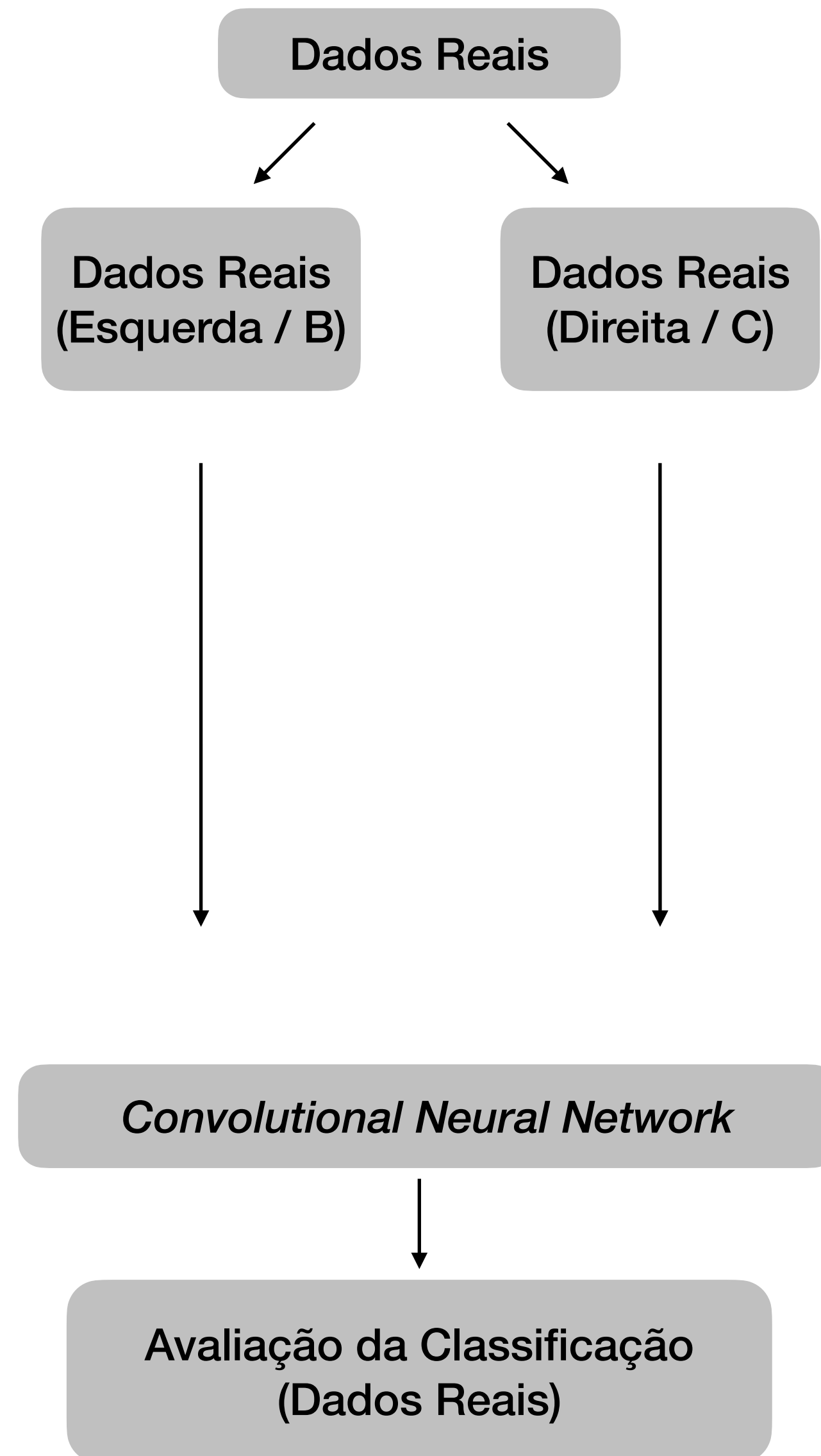


- A célula A12 possui 9 telhas em cada um dos seus 3 tubos.
- O scan com a fonte de Cs funciona como um raio-x de cada célula.
- Durante a passagem da fonte radioativa, em um dos tubos, vemos 9 picos no seu sinal temporal.

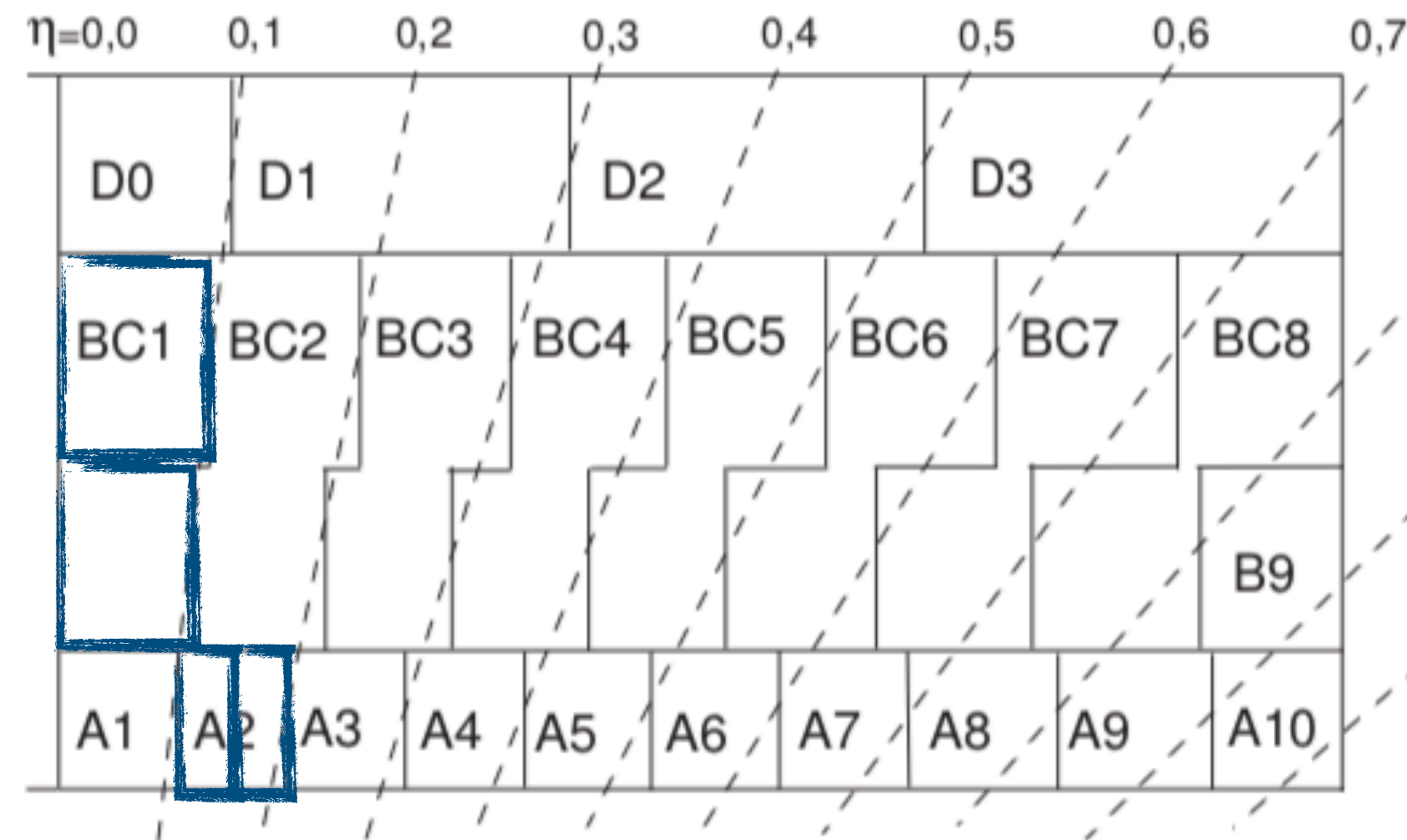
Roadmap



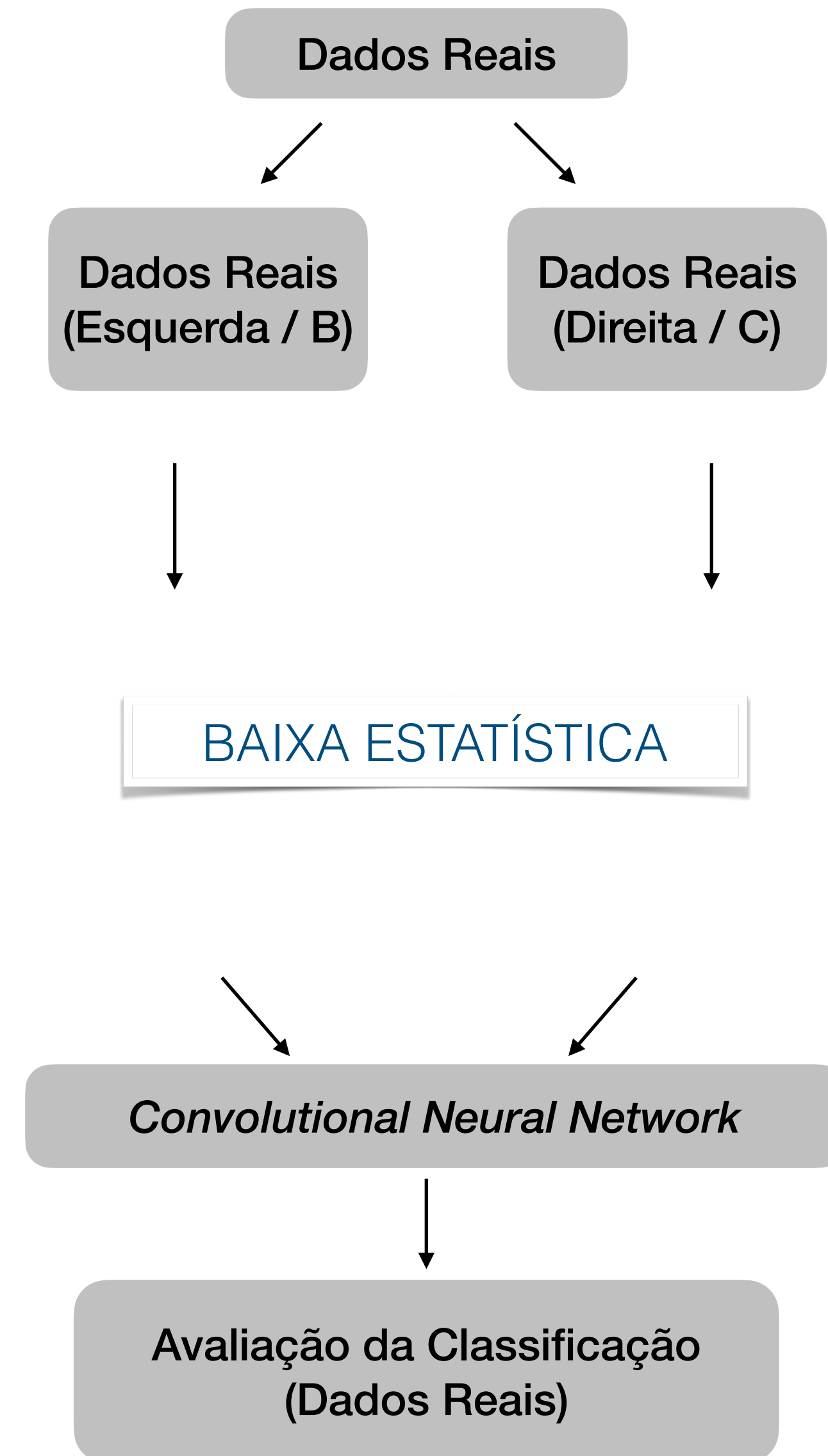
- Usando o sinal coletado pela PMT de anodo único podemos dividir as células A em duas novas regiões: esquerda e direita.
- Com um método similar as células BC são divididas em células B e C, separadamente. Desta vez usando informações dos tubos presentes em cada uma destas células.
- Essa nova separação permite criar um banco de dados anotado e tratar o problema sob a perspectiva do **aprendizado supervisionado**.



Roadmap

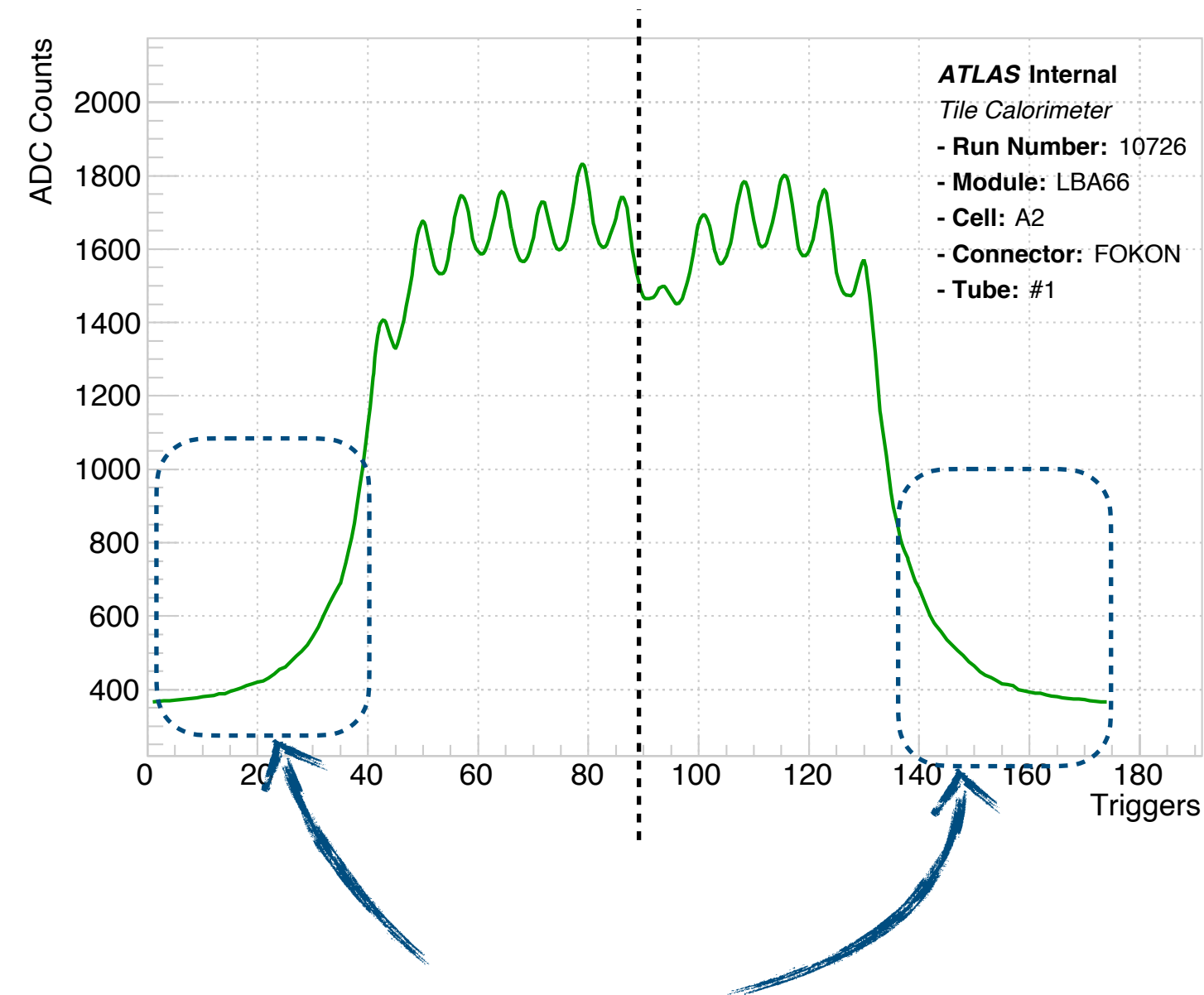


- Usando o sinal coletado pela PMT de anodo único podemos dividir as células A em duas novas regiões: esquerda e direita.
- Com um método similar as células BC são divididas em células B e C, separadamente. Desta vez usando informações dos tubos presentes em cada uma destas células.
- Essa nova separação permite criar um banco de dados anotado e tratar o problema sob a perspectiva do **aprendizado supervisionado**.



Roadmap

Sinal coletado pela PMT



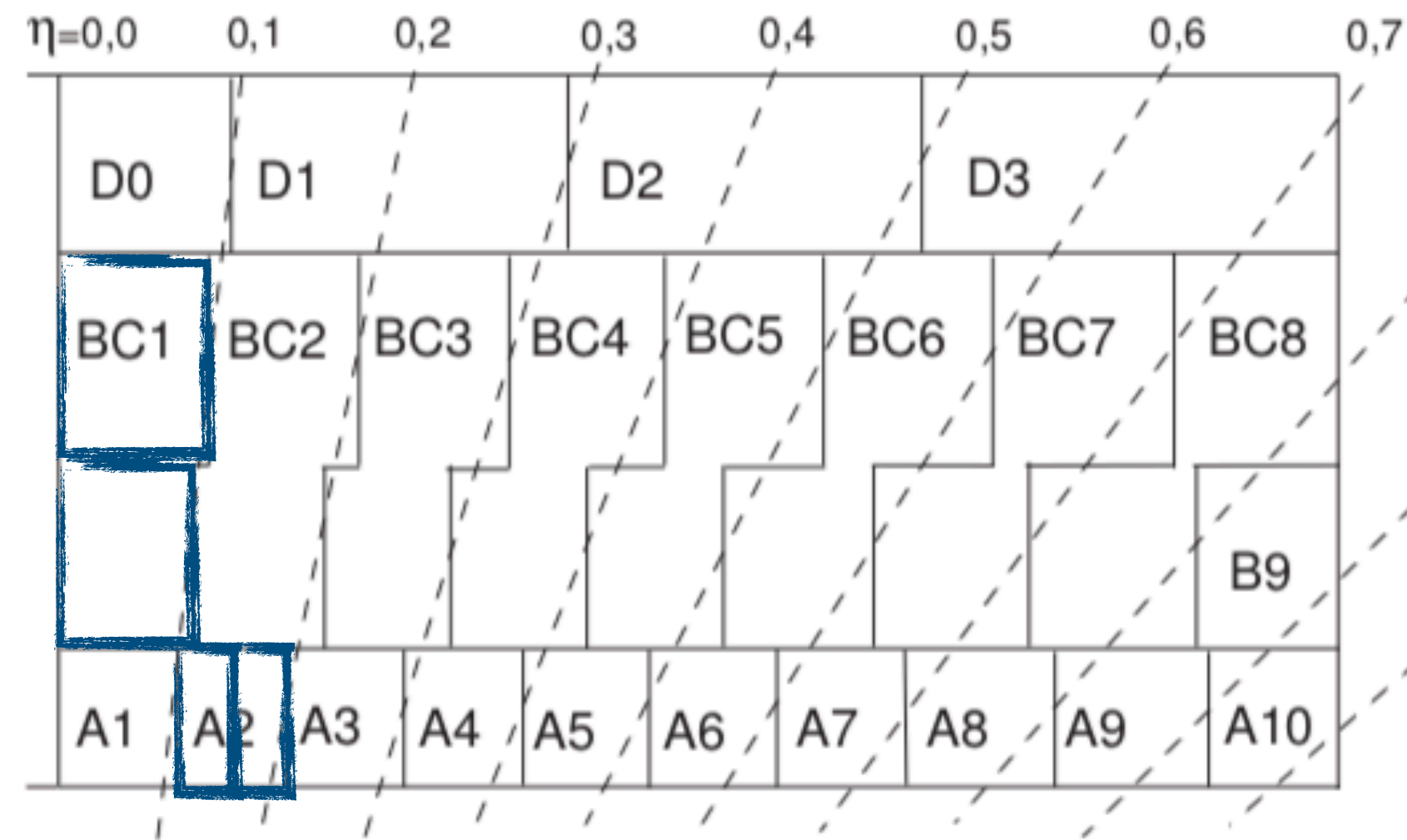
Não descartamos os sinais de baixa amplitude

A2 Cs Run Numbers

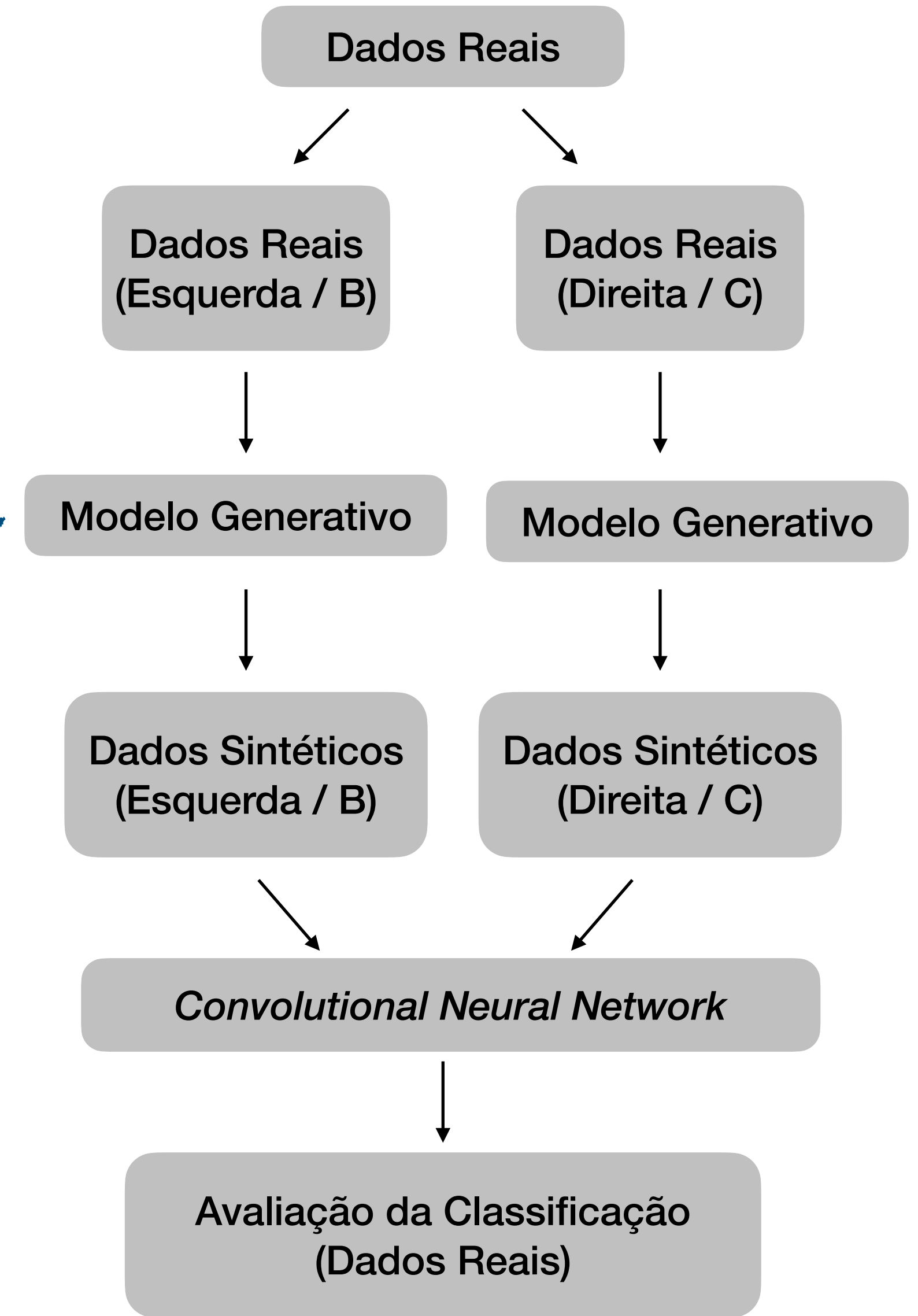
Amostras

10726F	520
10727B	524
10728F	524
10729B	528
10730F	524
10731B	524
10732F	524
10733B	524
10734F	522
10735B	522
10736F	520
10737B	520
10741F	516
10742B	518
10743F	520
10744B	514

Roadmap



Gerar **imagens sintéticas** como as imagens formadas na grid de pixels da MA-PMT pode resolver o problema da baixa estatística coletada durante o processo de calibração.



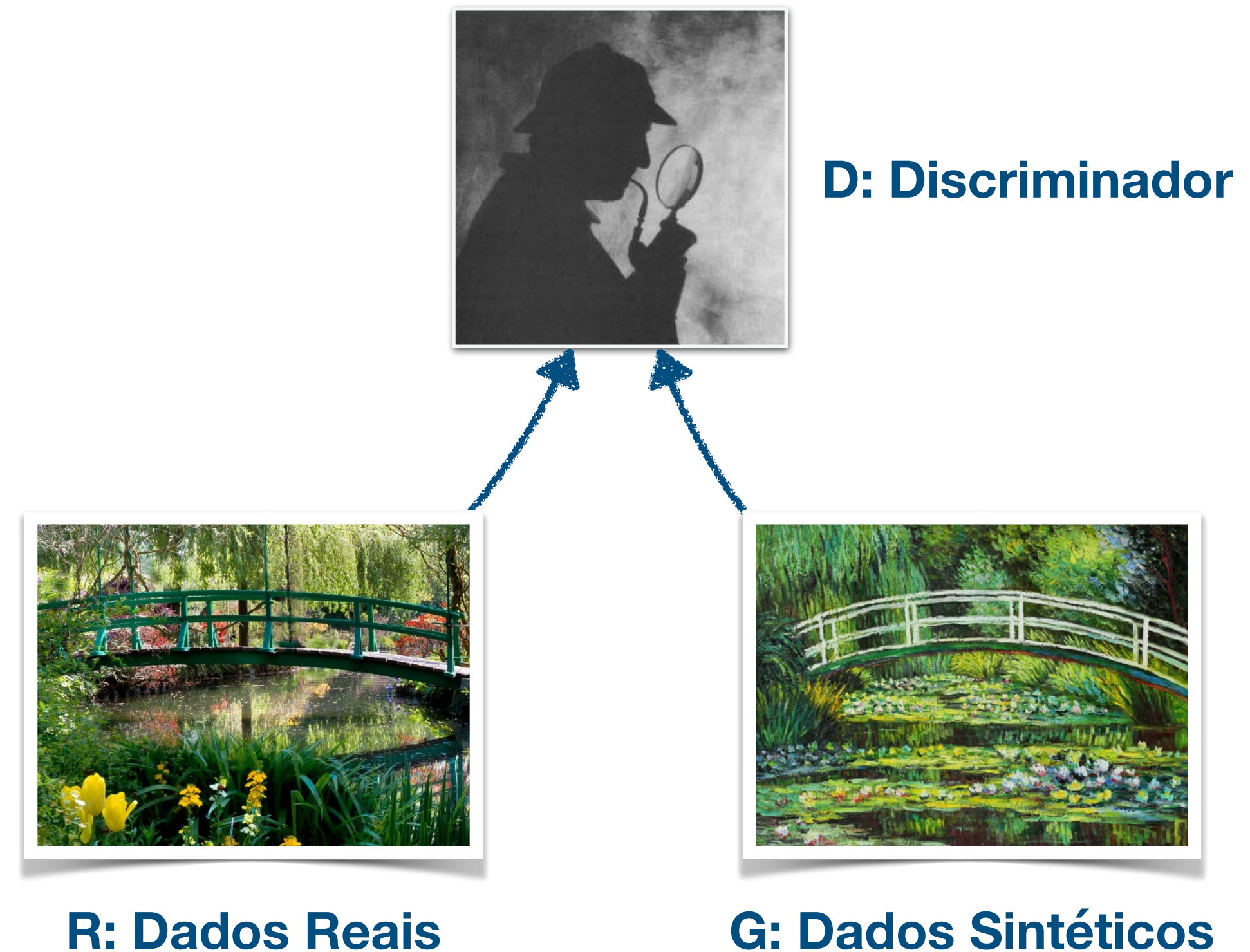
Modelos Generativos

- Usar um model generativo para aumentar o número de amostras no conjunto de dados.
- **Generative Adversarial Networks** (GAN) são redes baseadas em teoria dos jogos.
- Dois modelos (*Deep Neural Networks*) competem entre si.
- **Gerador:** Cria imagens sintéticas que se parecem com as imagens reais.
- **Discriminador:** Avalia se as images são reais ou sintéticas.

MinMax Game

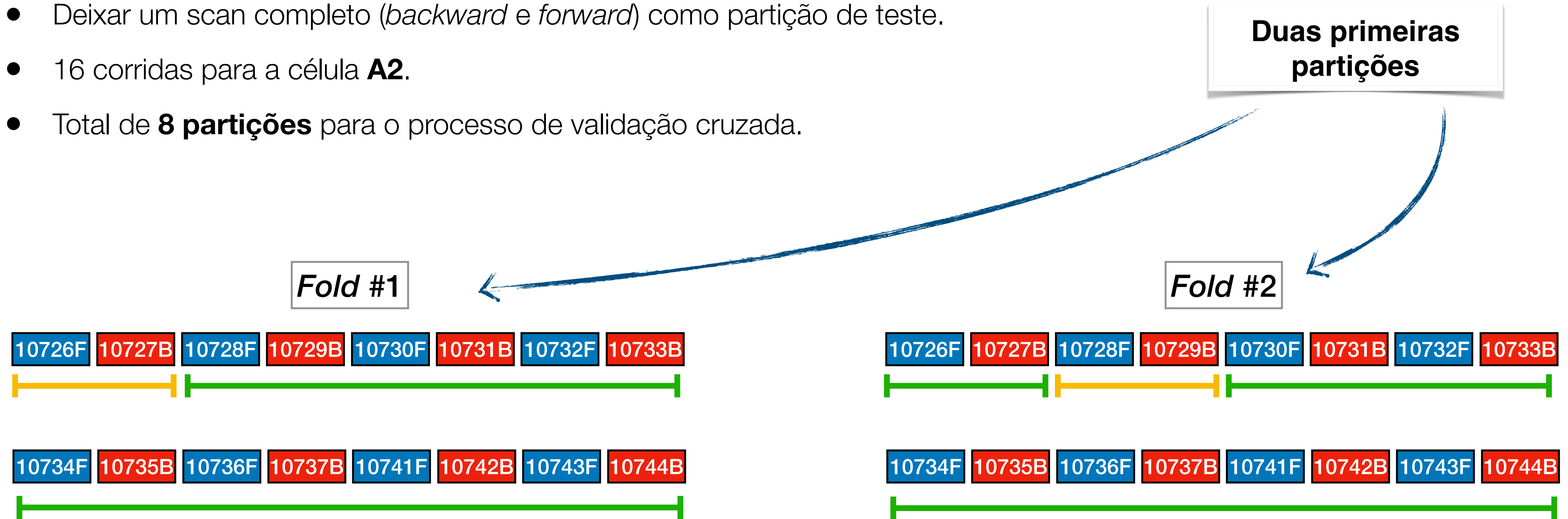
$$J^{(D)} = -\frac{1}{2}\mathbb{E}_{x \sim p_{data}} \log D(x) - \frac{1}{2}\mathbb{E}_z \log(1 - D(G(z)))$$

$$J^{(G)} = -J^{(D)}$$



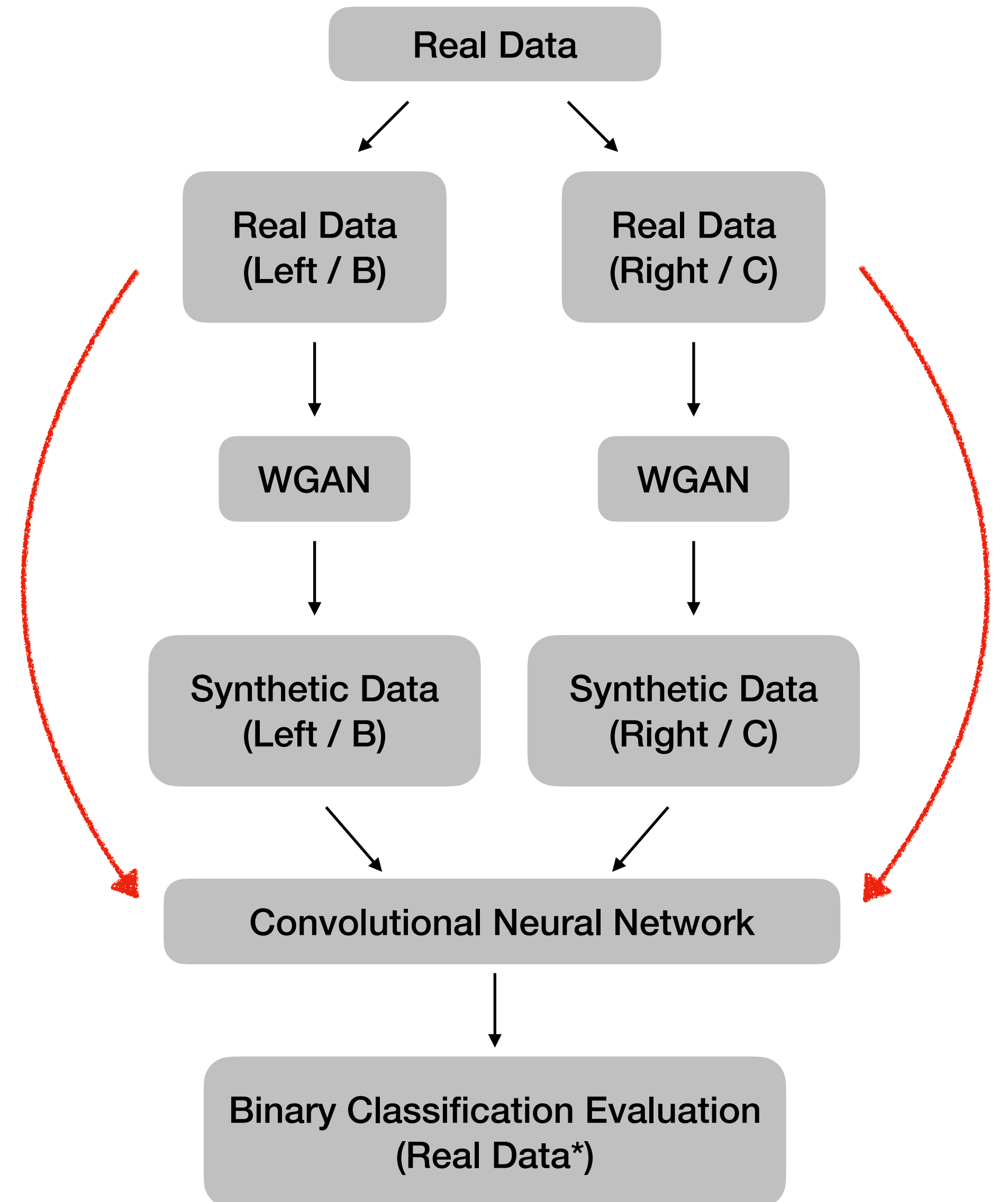
Proposta de Validação Cruzada

- Deixar um scan completo (*backward* e *forward*) como partição de teste.
- 16 corridas para a célula **A2**.
- Total de **8 partições** para o processo de validação cruzada.



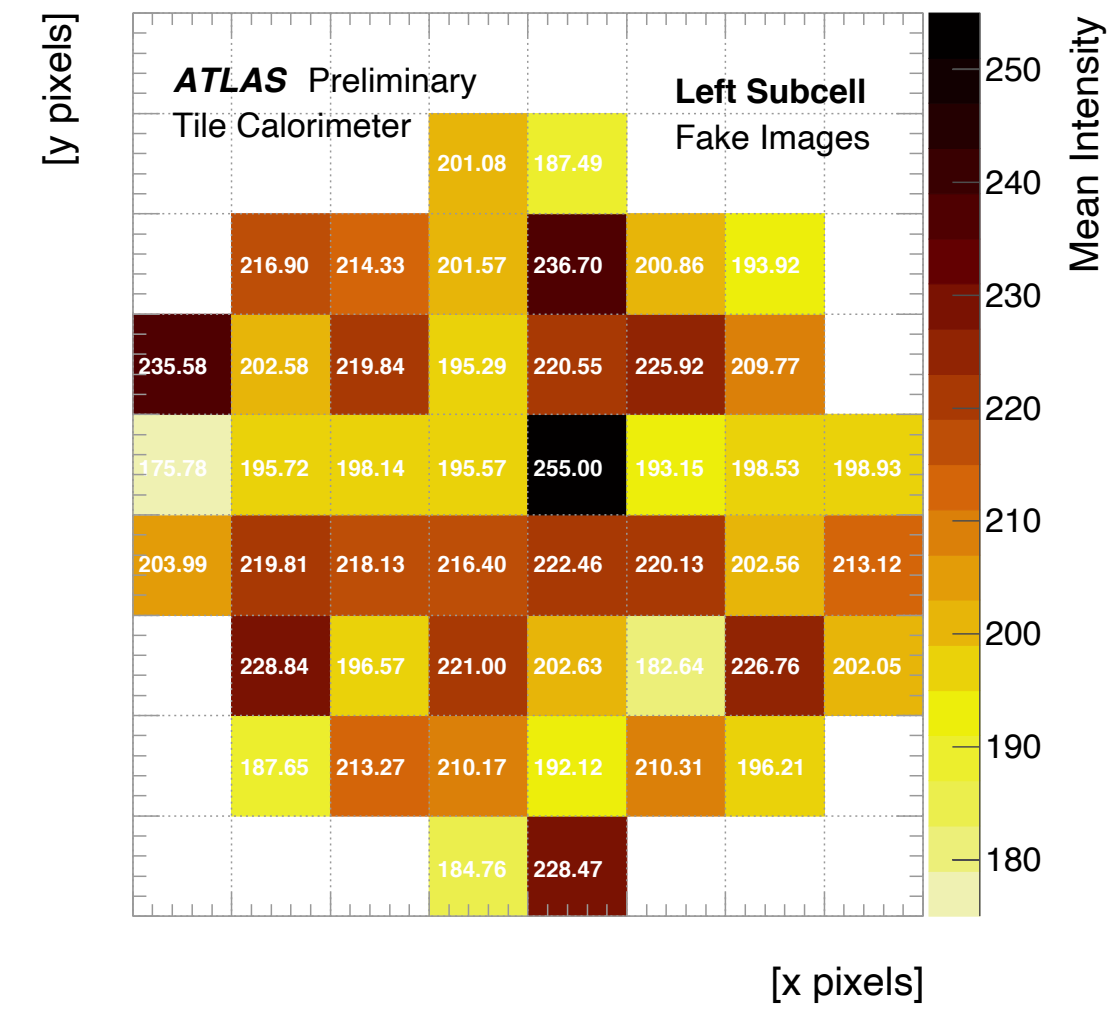
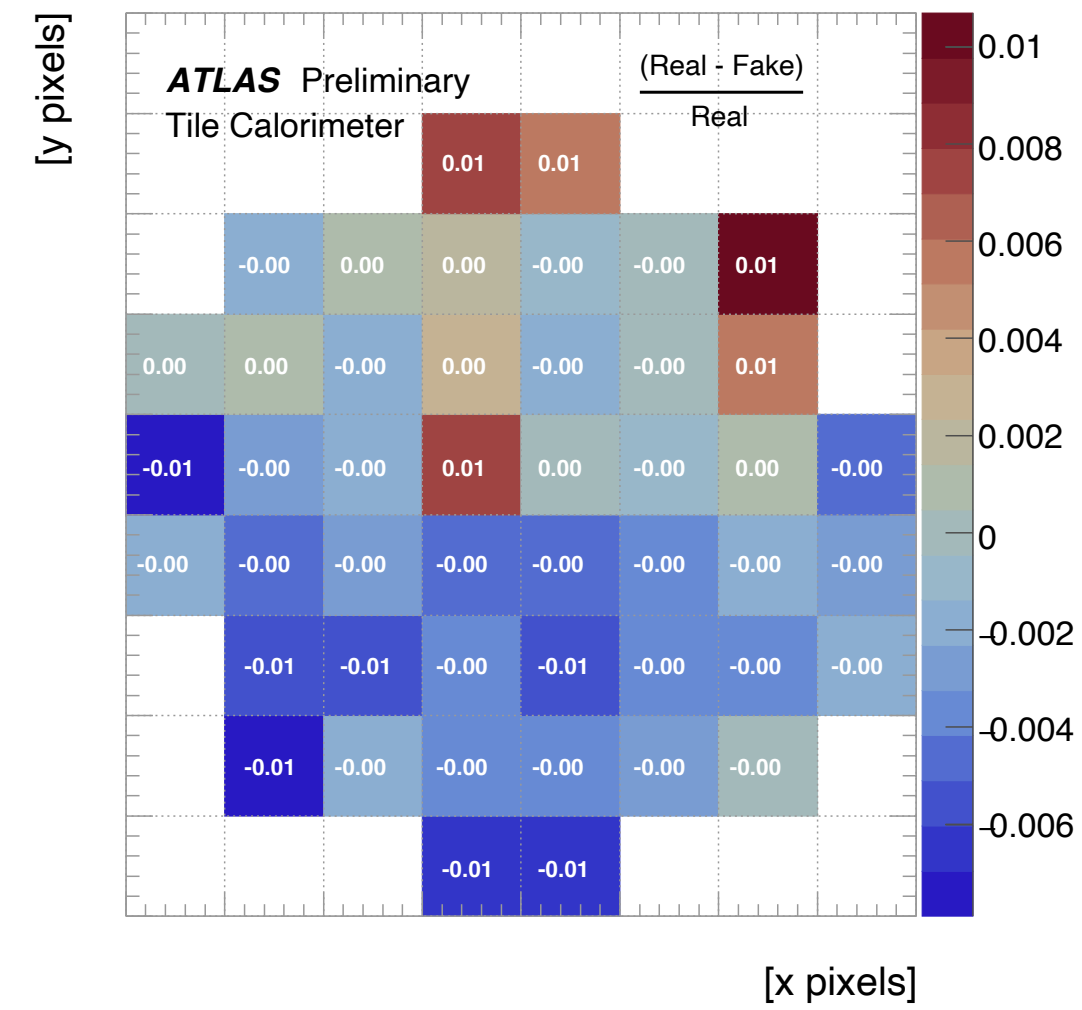
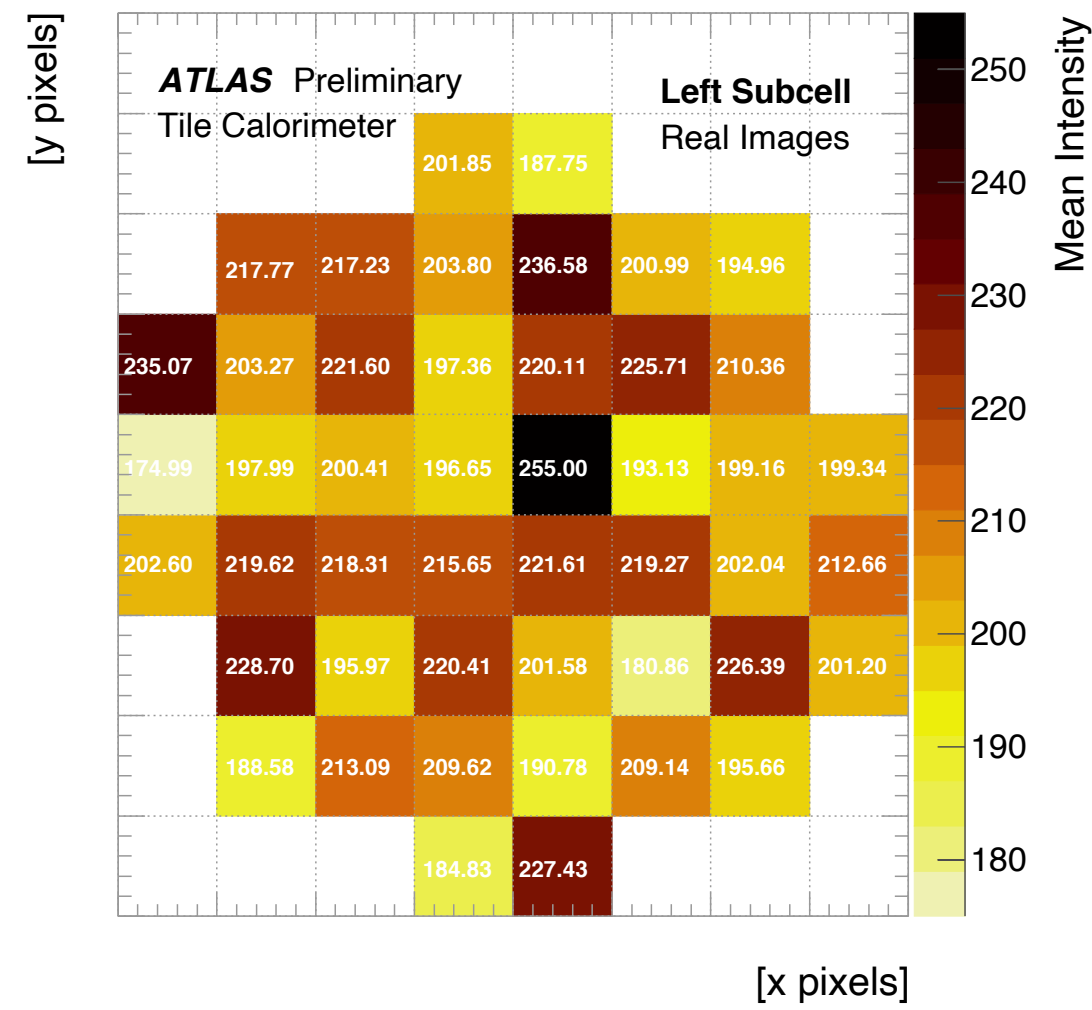
Pipeline de Dados

- Propomos três *pipelines* diferentes.
- **Synthetic:** Treinar a CNN com dados sintéticos.
- **Altogether:** Treinar a CNN com uma combinação de dados reais e sintéticos.
- **Fine Tuning:** Carregar a CNN treinada com dados sintéticos, e começar um processo de ajuste fino do modelo usando os dados reais.

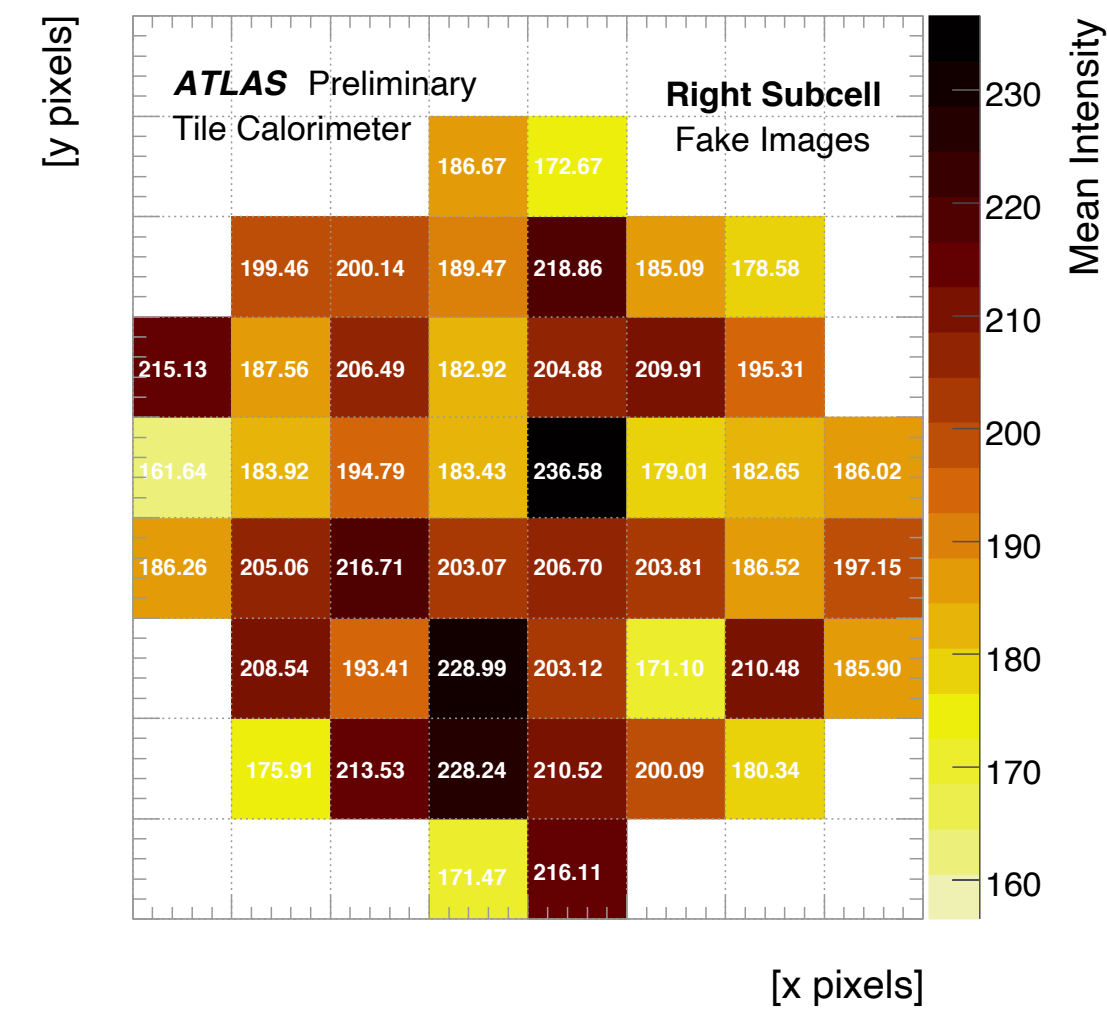
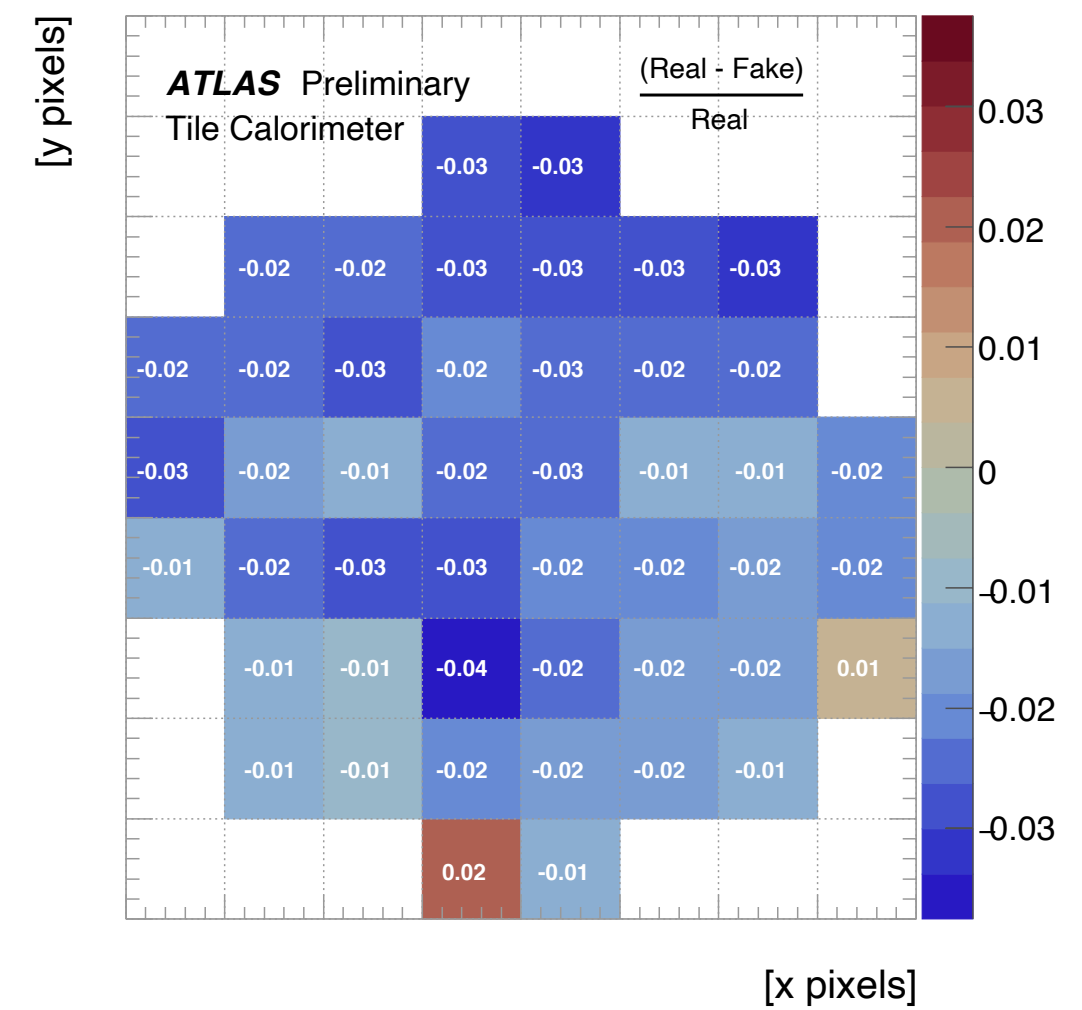
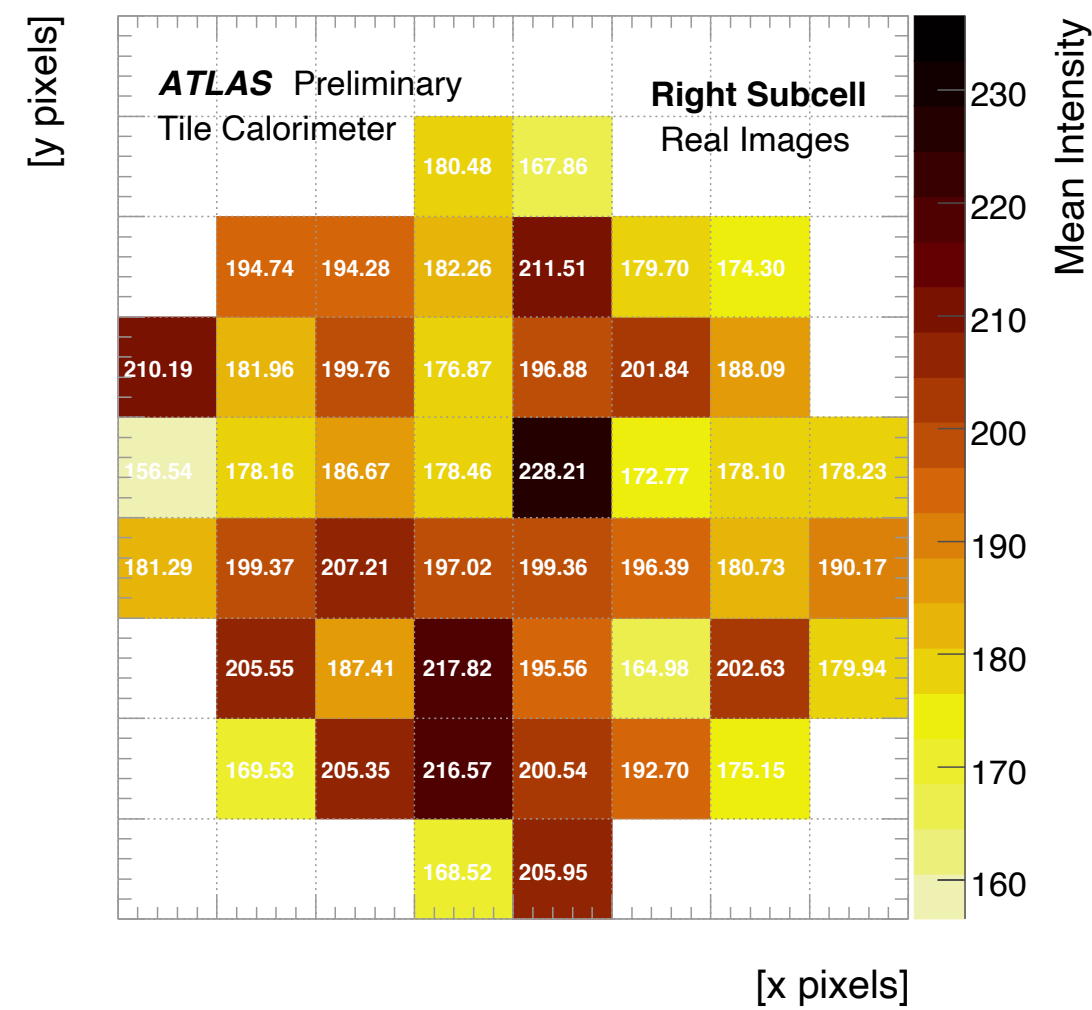


Resultados

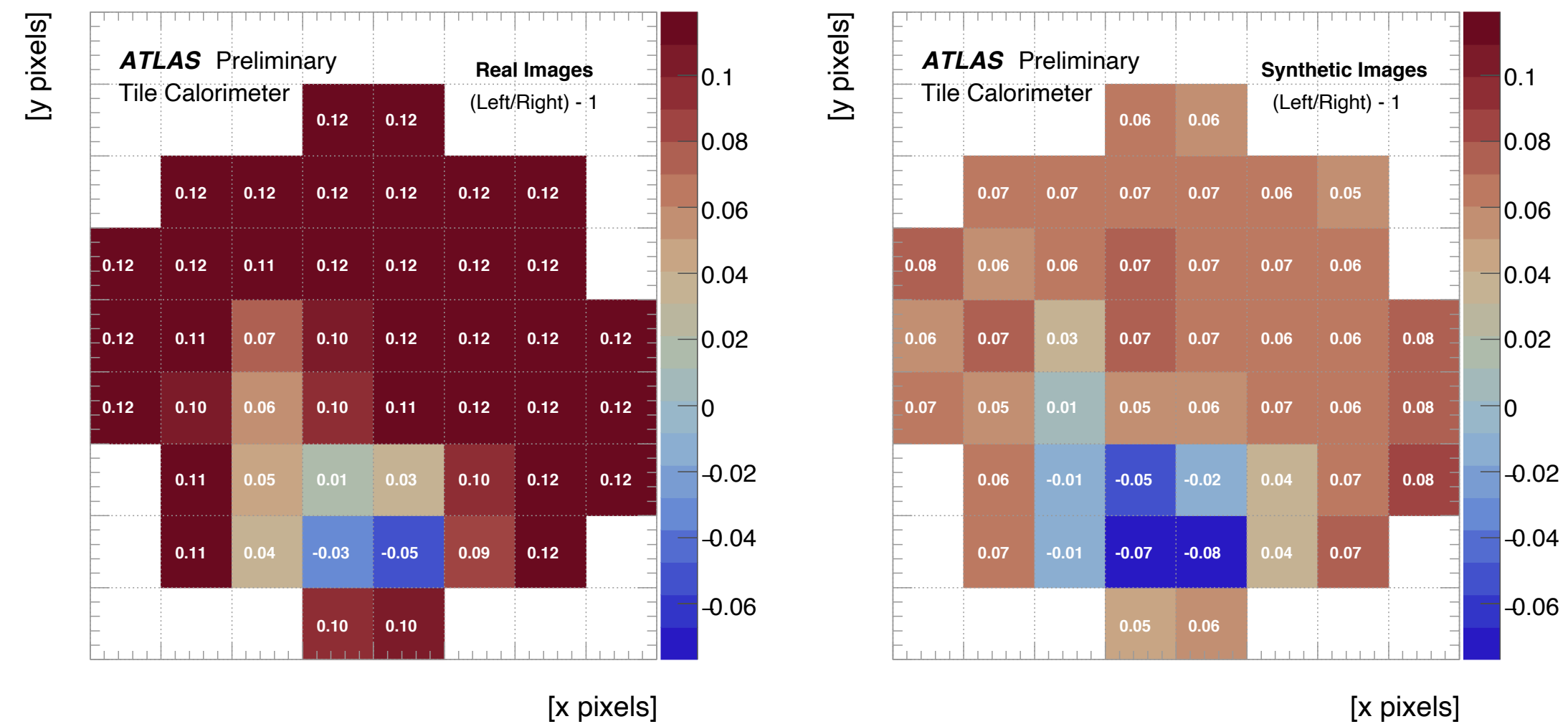
Imagens - Célula A2



- Exemplo de imagens reais e sintéticas geradas para a célula A2.
- As imagens mostram a média de intensidade de pixels para cada conjunto de dados.
- Diferença relativa entre as imagens reais e sintéticas possui 3% como o maior valor.



Imagens - Célula A2

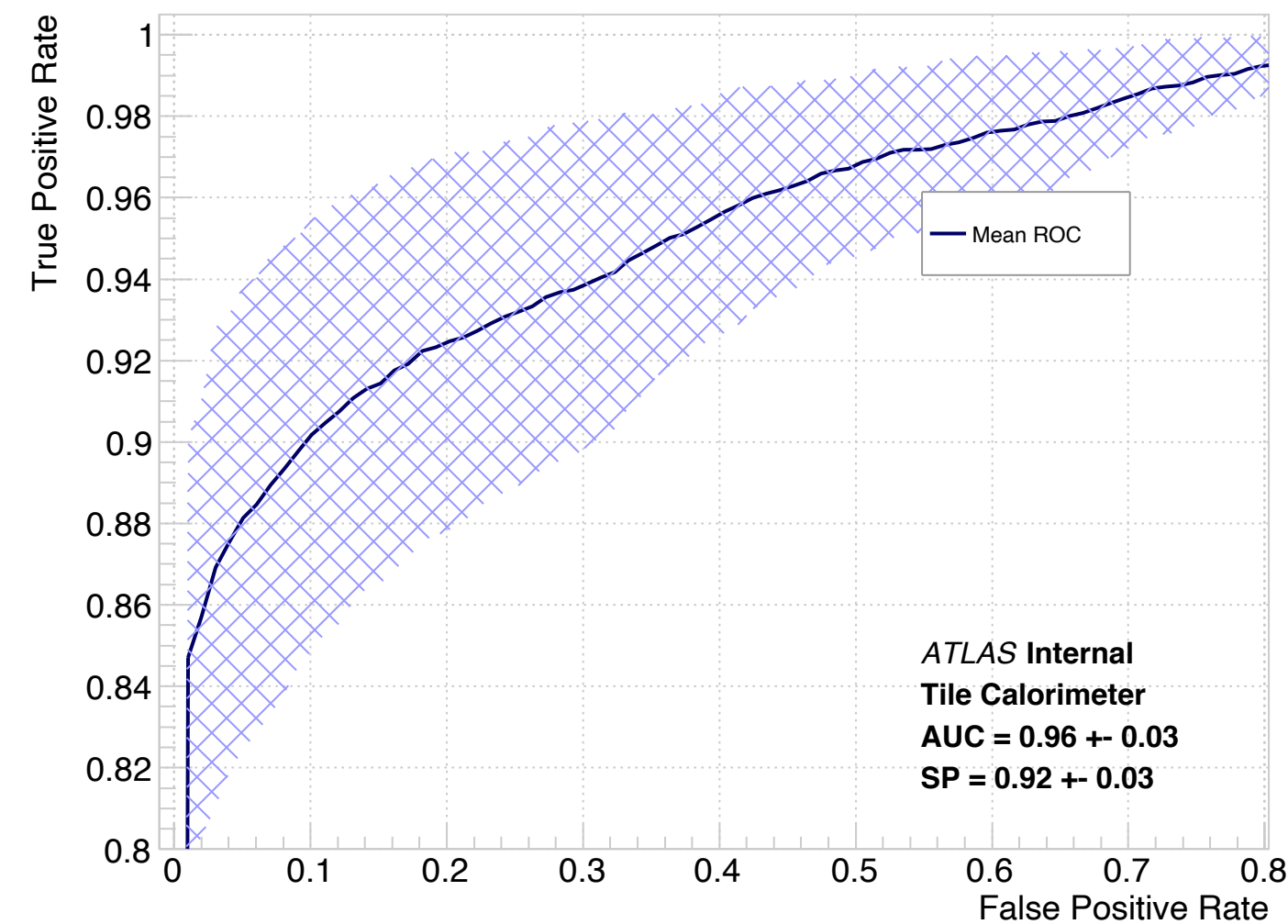


Imagens sintéticas foram capazes de capturar o padrão que difere cada uma das classes.

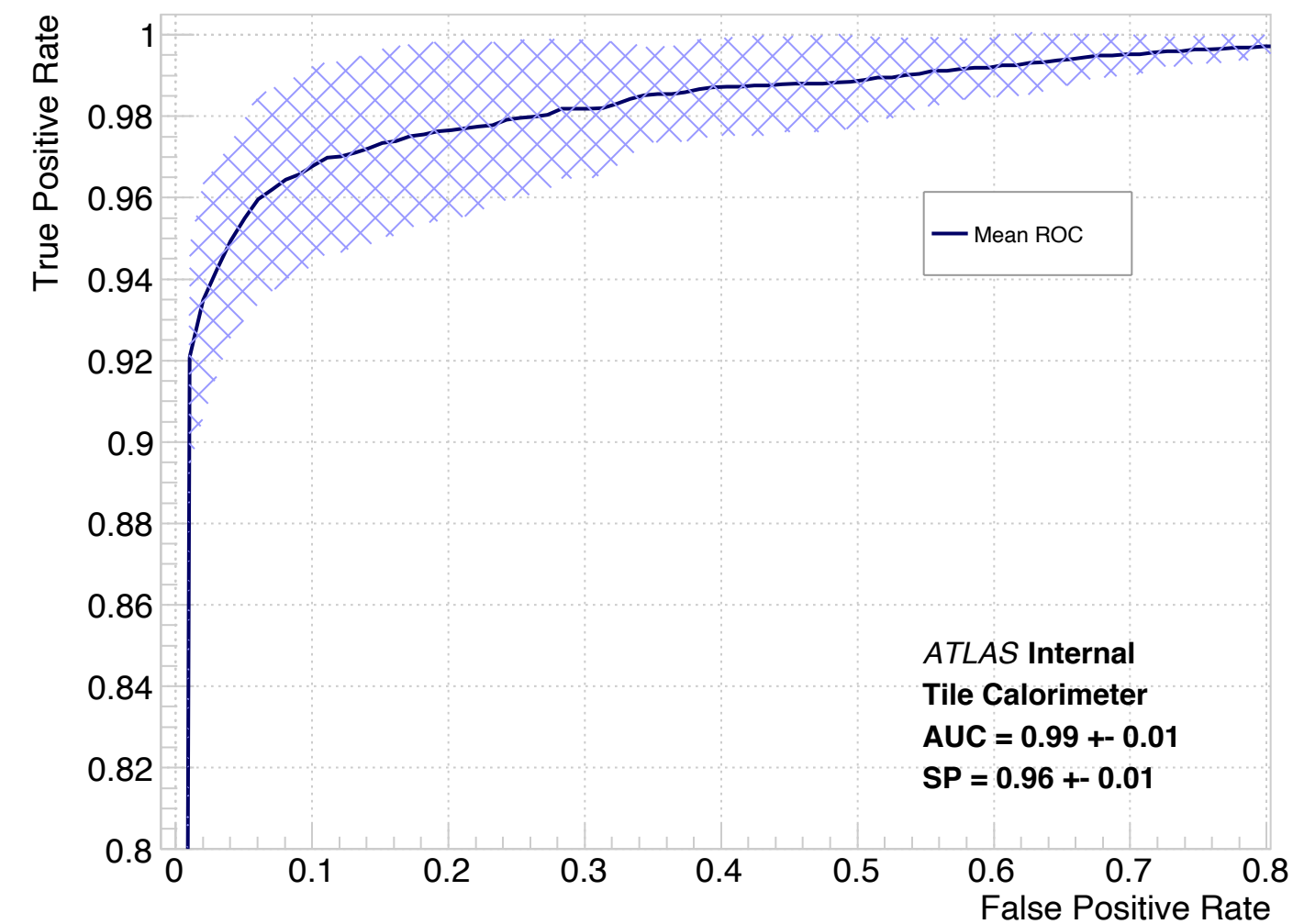
- Aqui mostramos a diferença no padrão discriminatório que representa cada uma das classes (sub-células).
- Para a camada (granularidade transversal) as classes são para as sub-células direita e esquerda.
- Na primeira figura vemos a diferença nesse padrão para as imagens reais (coletadas durante a calibração de Cs).
- A segunda figura mostra o mesmo padrão, desta vez para as imagens sintéticas geradas pelo model generativo.
- Claramente podemos ver que as imagens sintéticas acompanham o mesmo padrão discriminatório das imagens reais.

Classificação - Célula A2

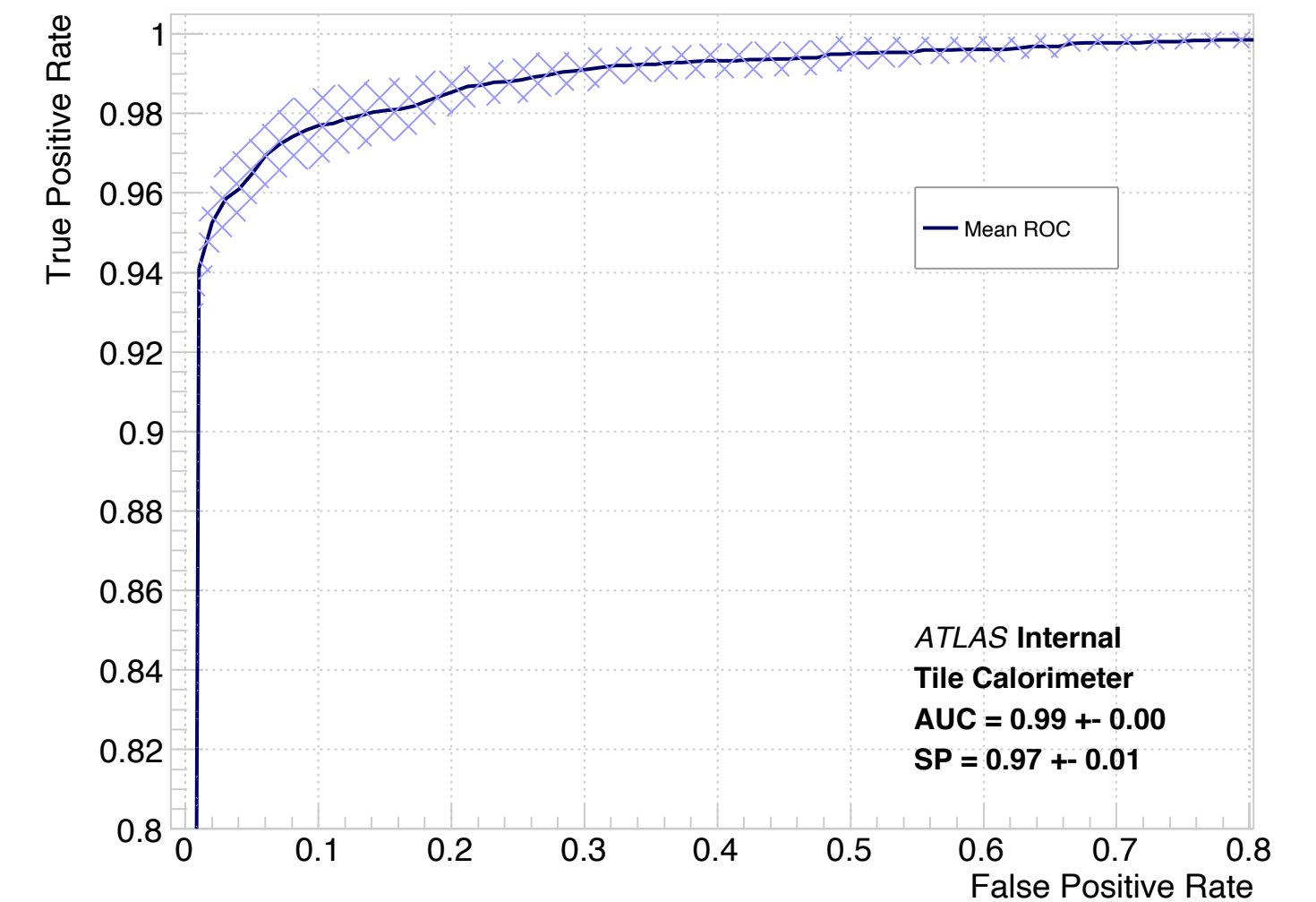
Synthetic



Altogether

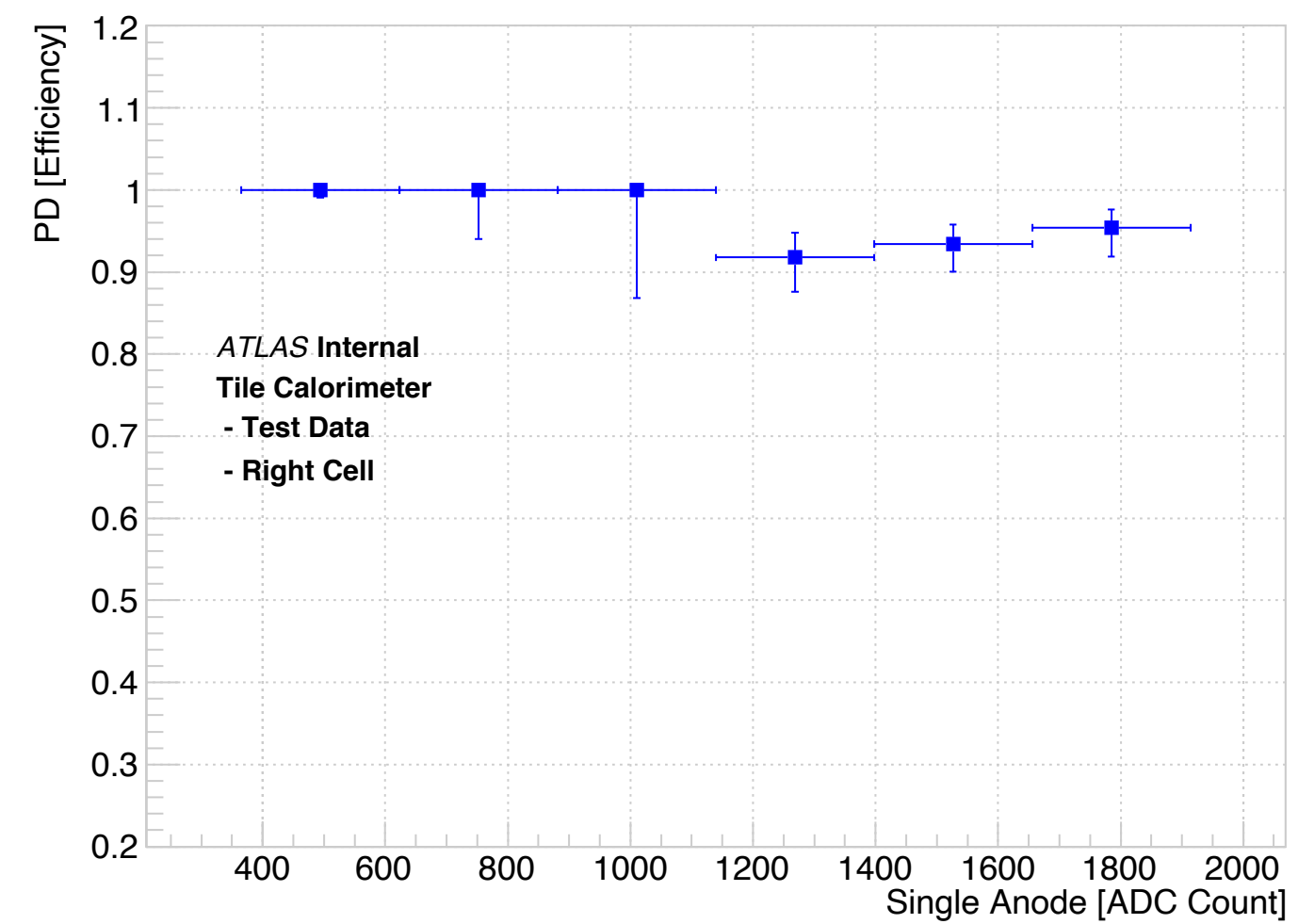
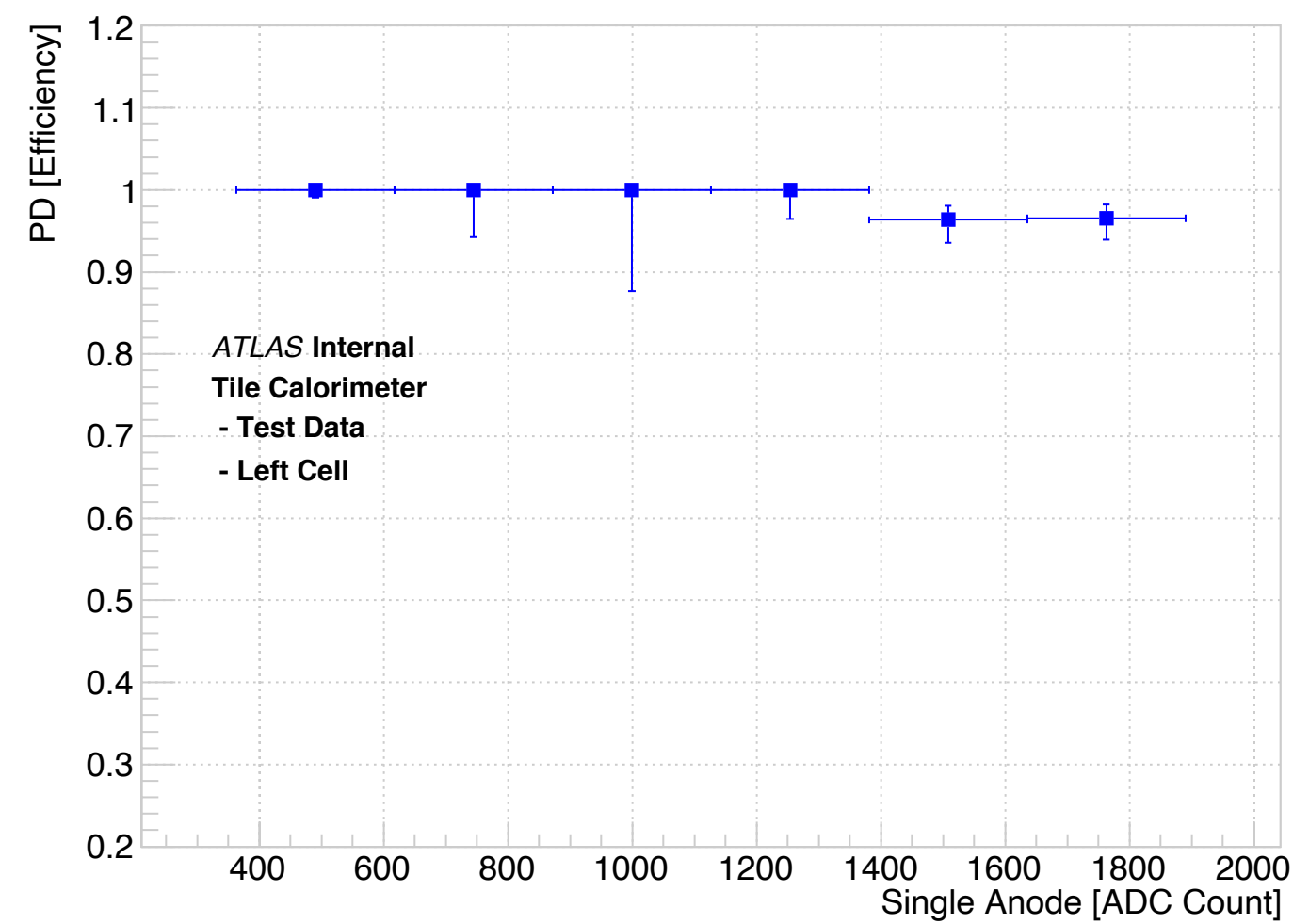


Fine Tunning



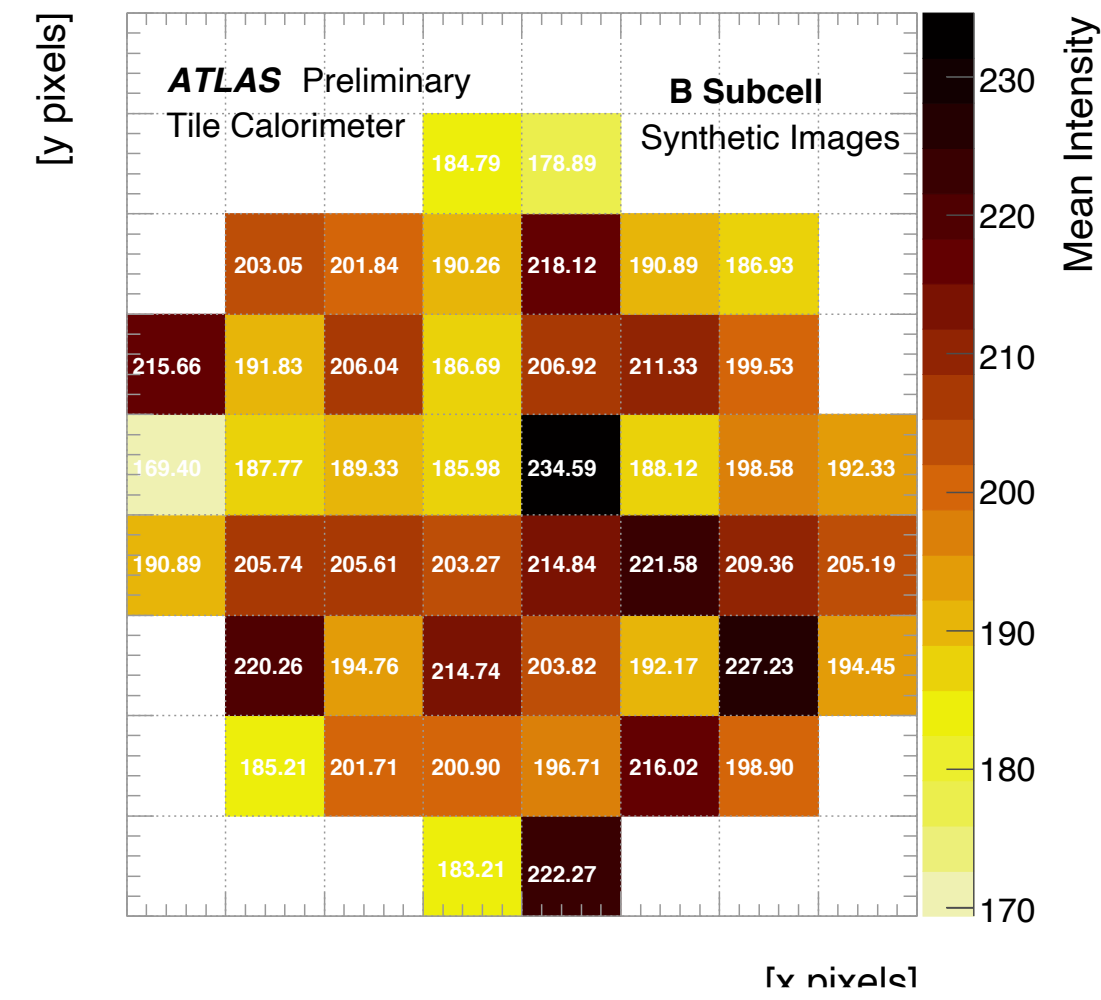
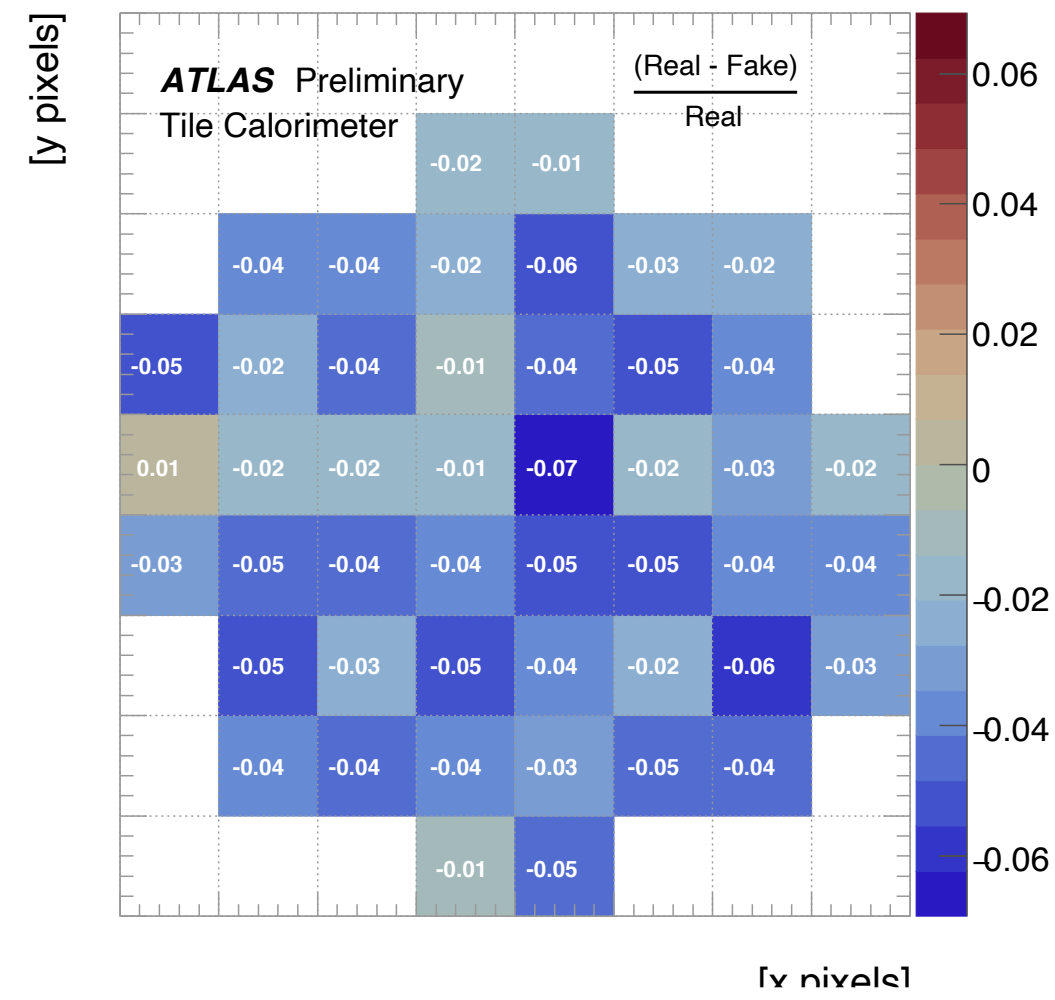
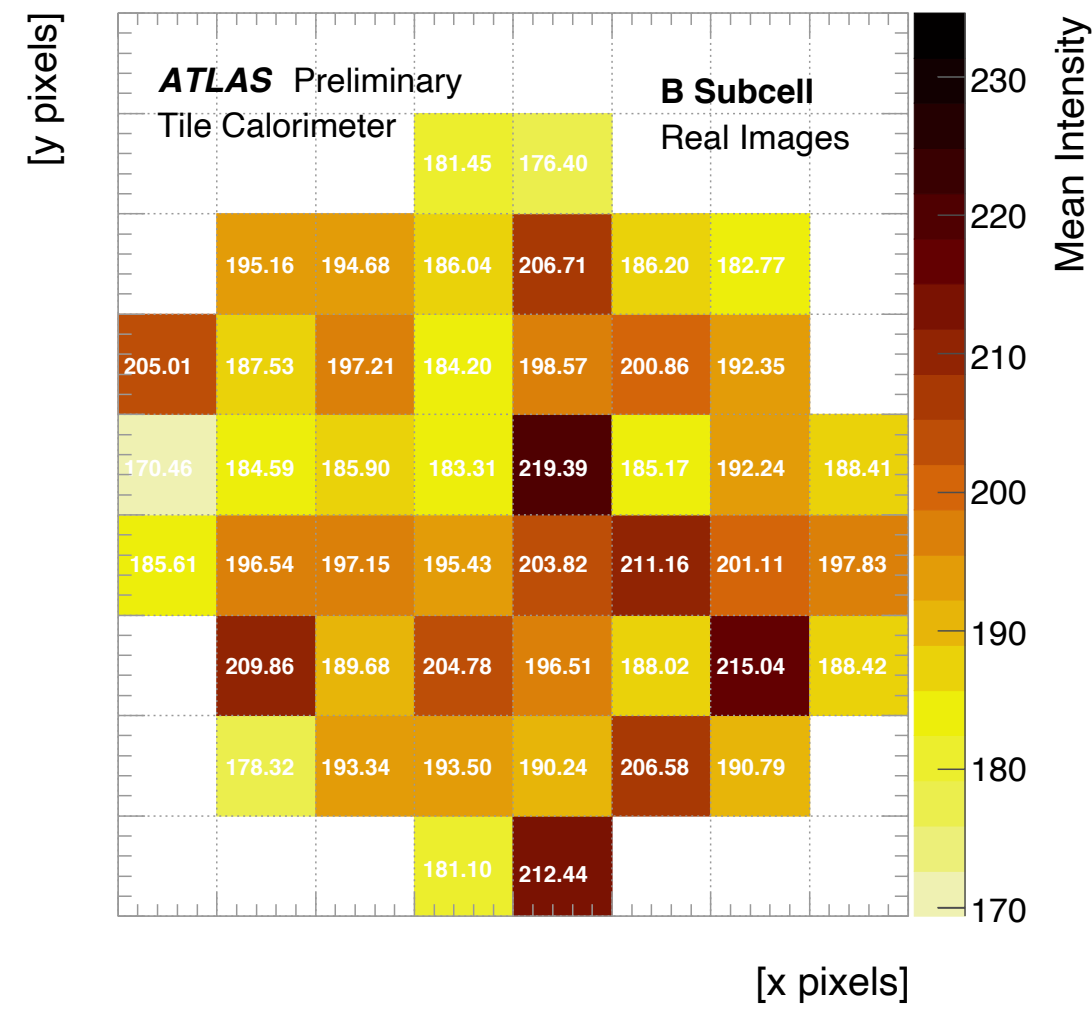
- As figuras acima mostram a curva ROC para diferentes pipelines de treinamento.
- O pipelines de treinamento possuem um impacto claro na estimativa de incerteza da classificação.
- O pipeline de *Fine Tuning* apresentou o melhor resultado entre os três pipelines propostos.

Classificação - Célula A2

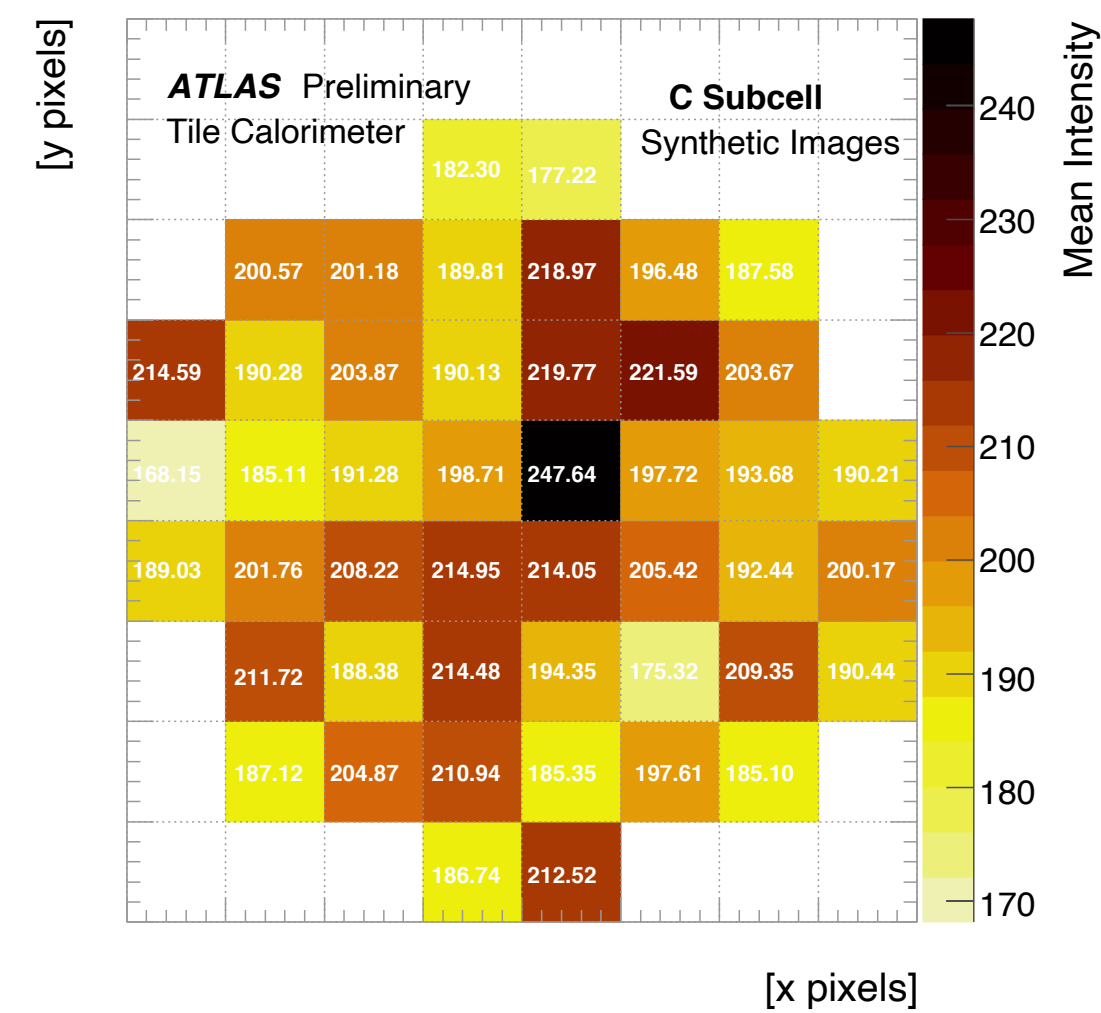
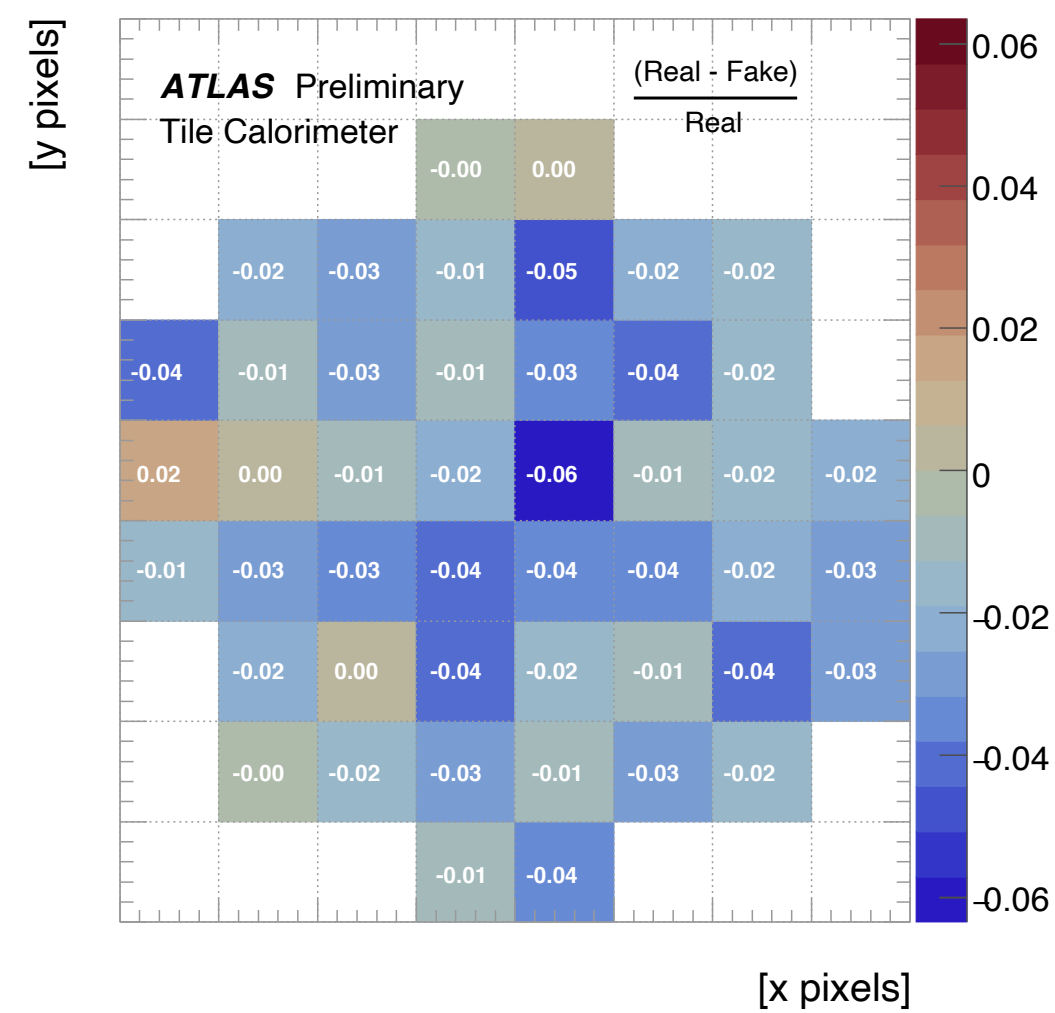
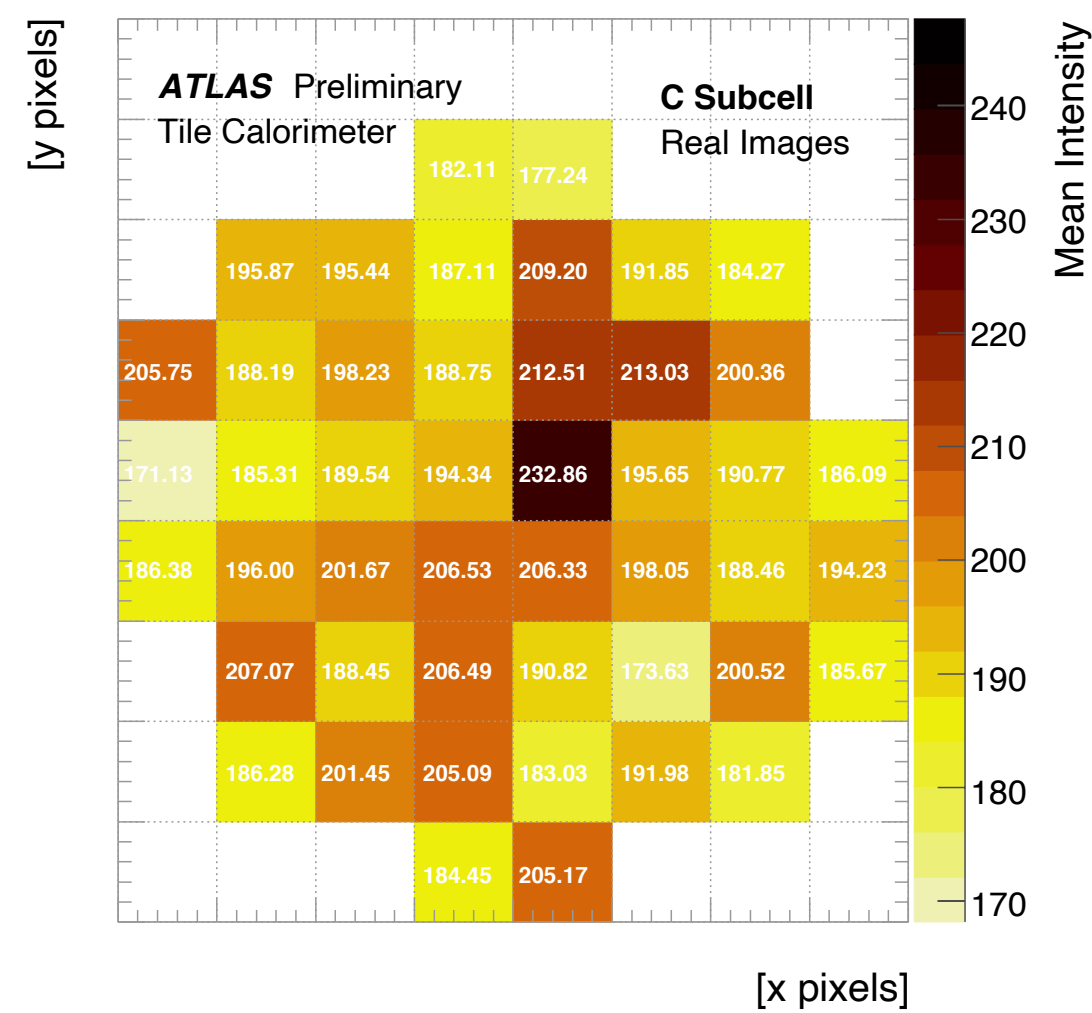


- Resultados na classificação da célula A2 mostram eficiência máxima em todos os bins, com exceção de alguns bins de alta amplitude.

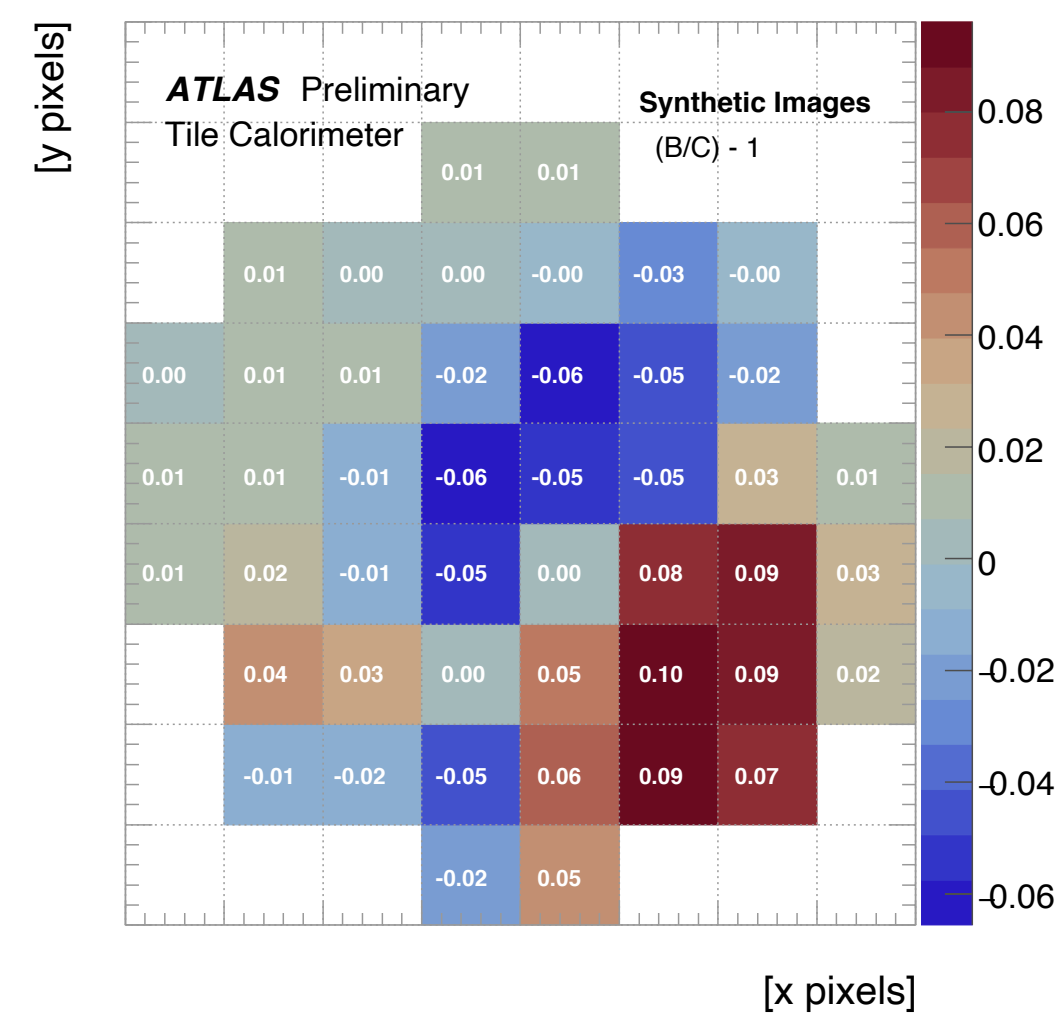
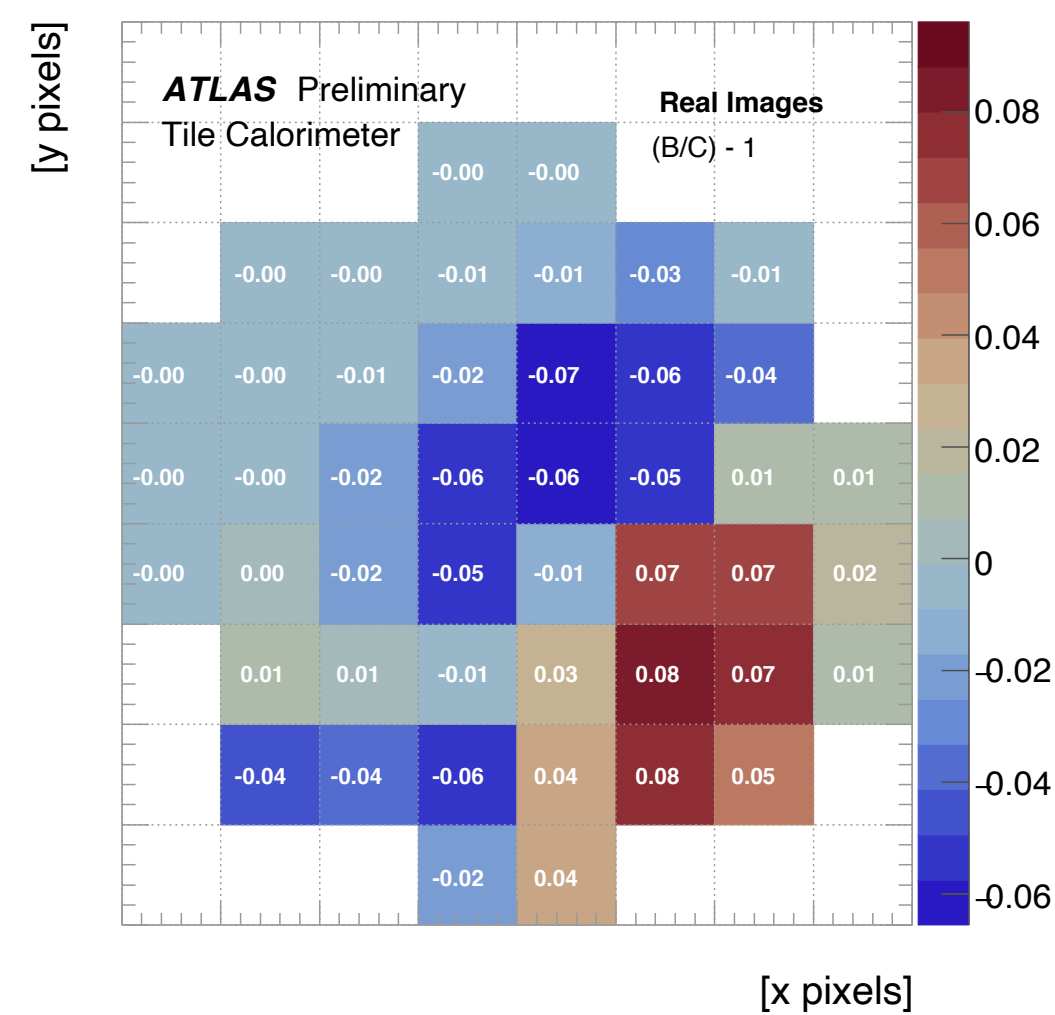
Imagens - Célula BC1



- Exemplos de imagens sintéticas geradas para a célula BC1.
- As imagens mostram a intensidade dia de pixels.
- Temos uma diferença relativa máxima de 6% em alguns pixels.



Imagens - Célula BC1

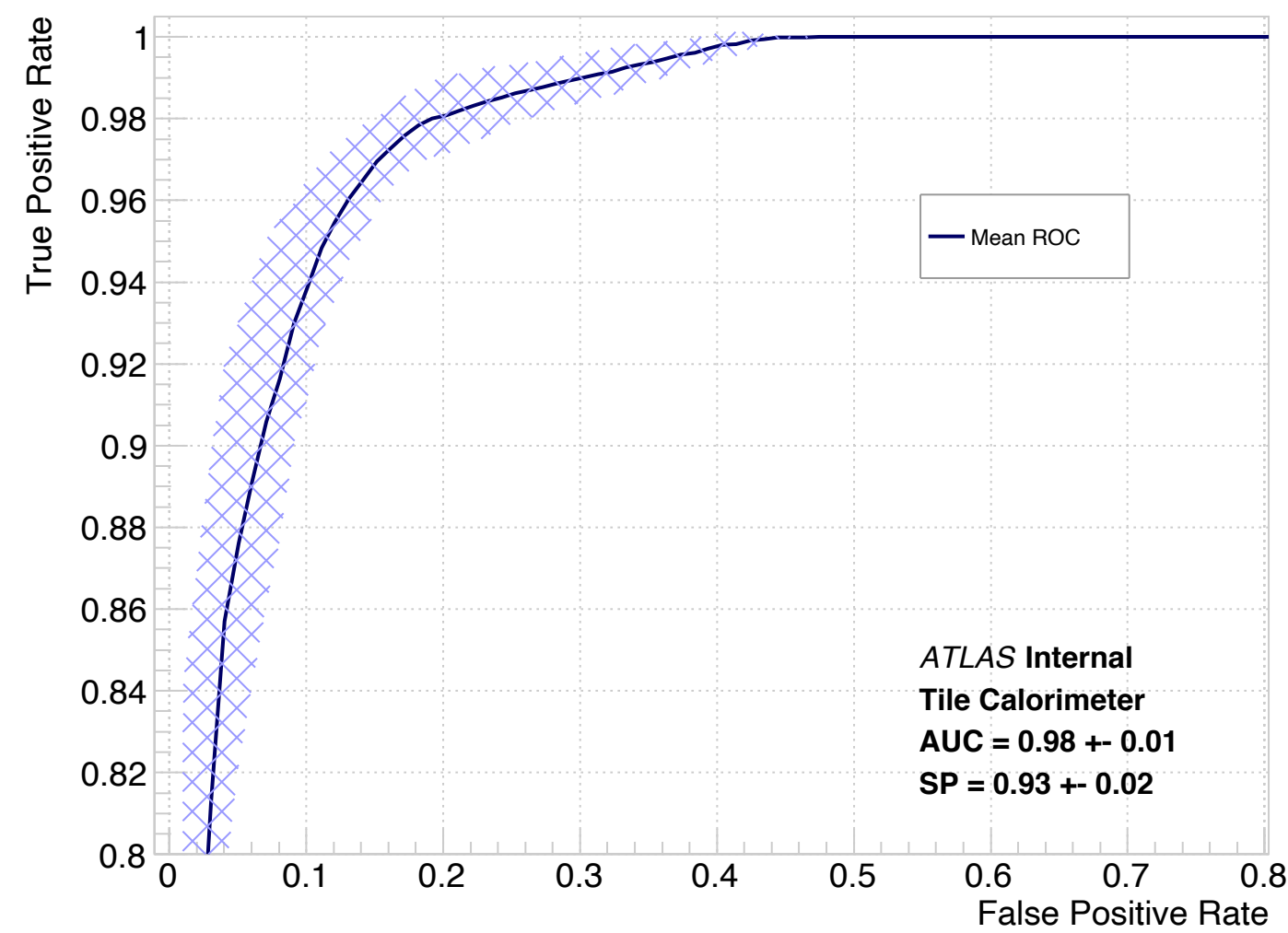


- Desta vez, as células para a camada BC, possuem classes de sub-células B e C, separadamente (granularidade longitudinal).
- Novamente, podemos observar que as imagens sintéticas conseguem capturar o mesmo padrão discriminatório visto nas imagens reais.

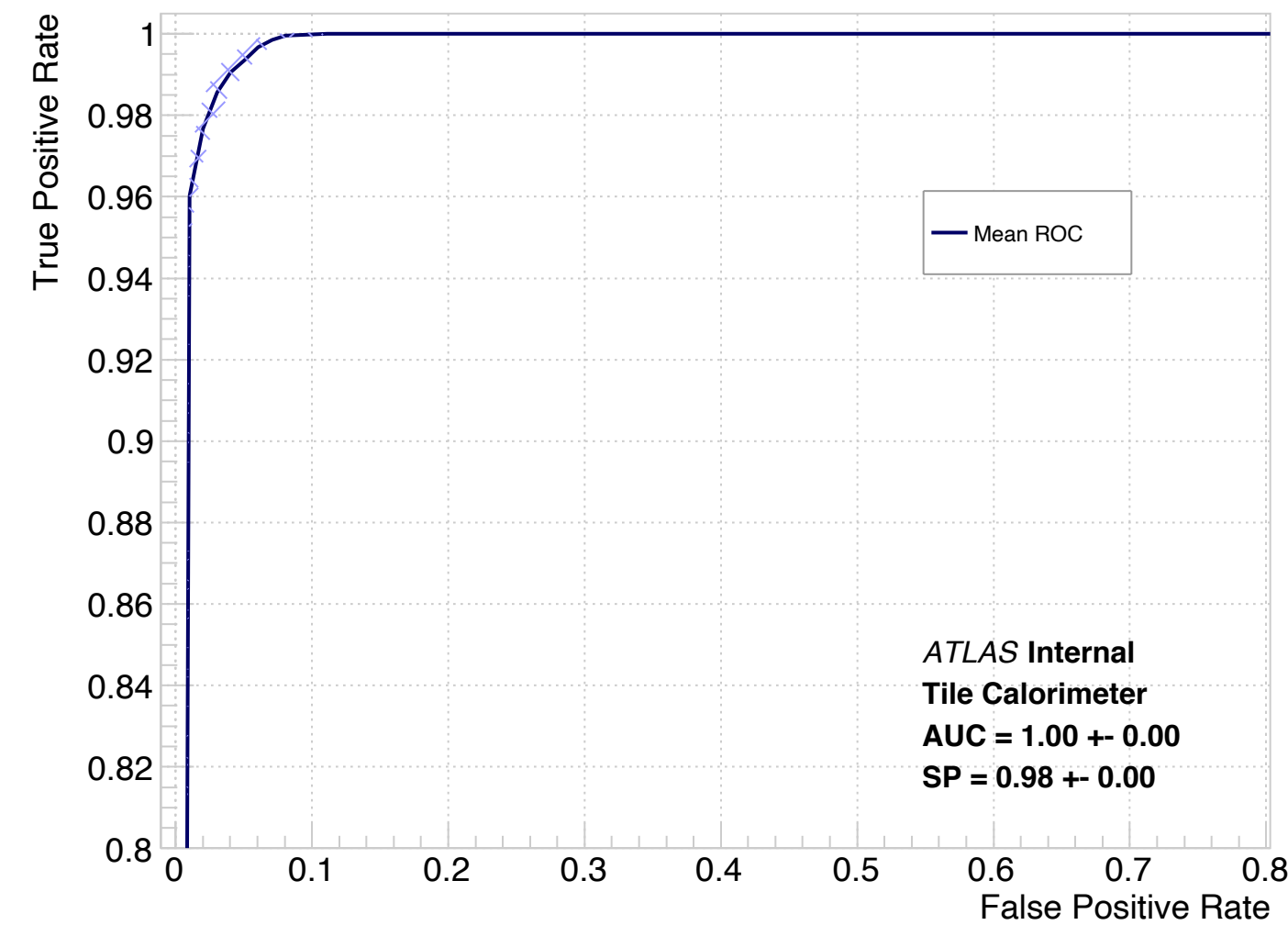
Imagens sintéticas foram capazes de capturar o padrão que difere cada uma das classes.

Classificação - Célula BC1

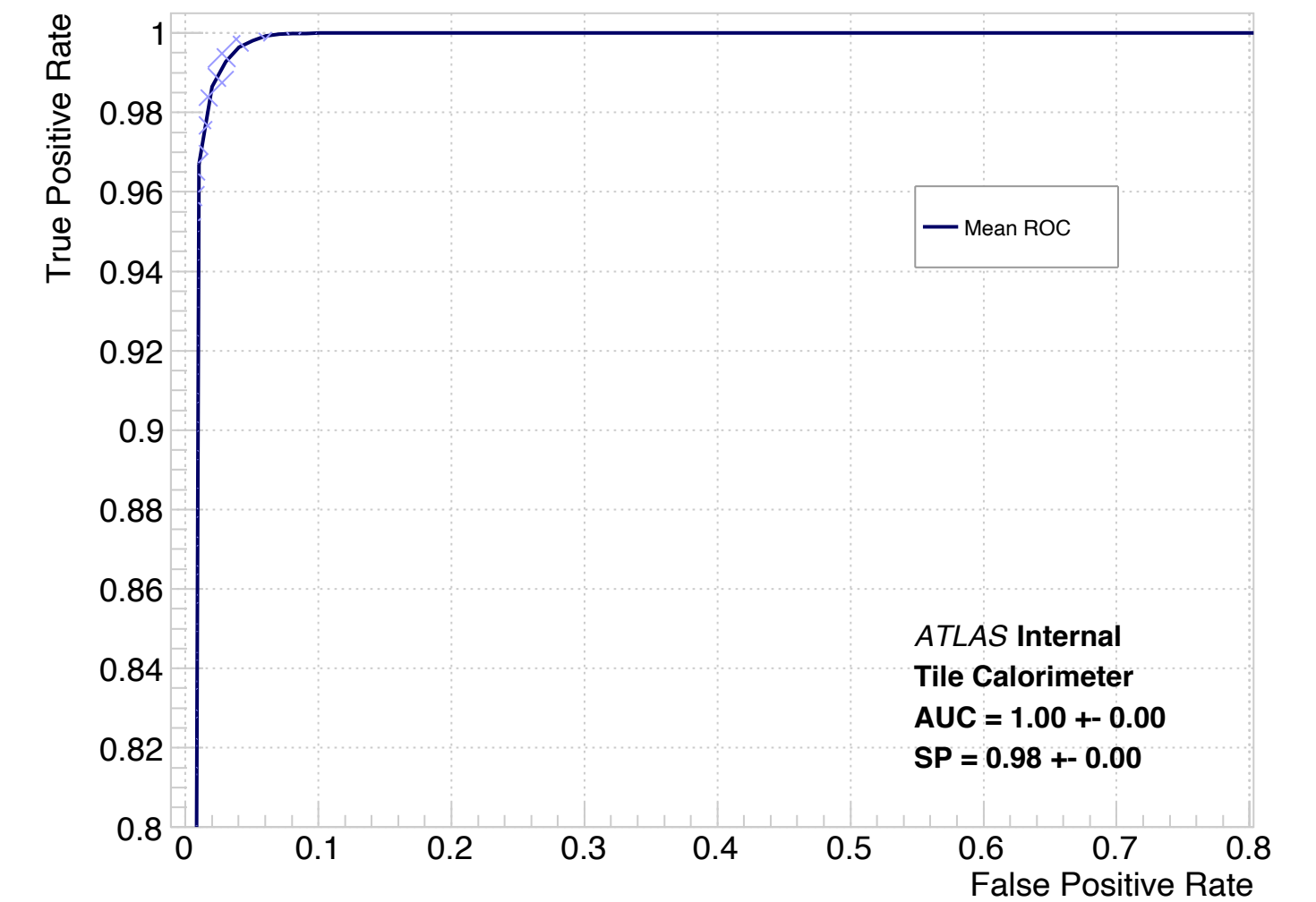
Synthetic



Altogether

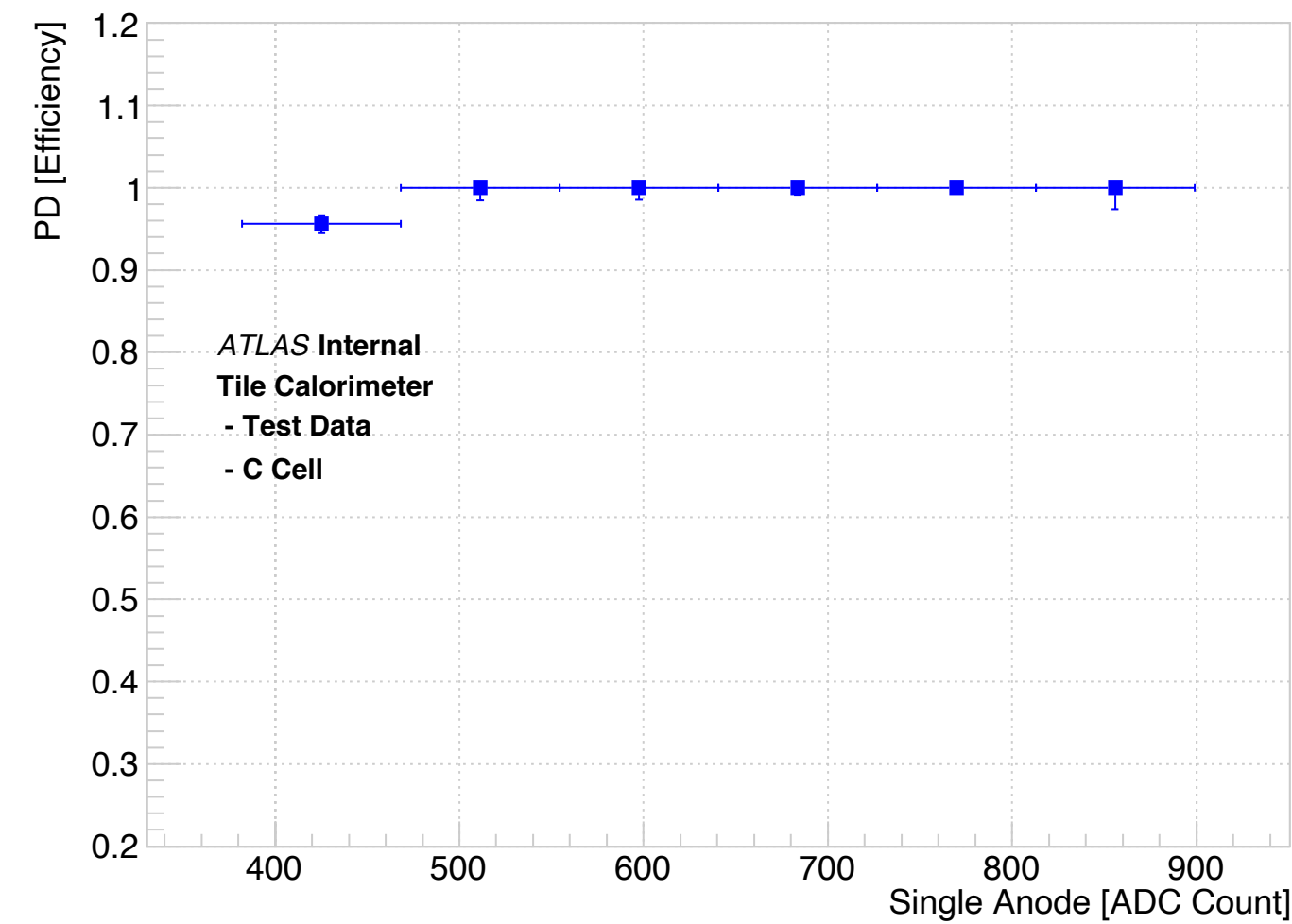
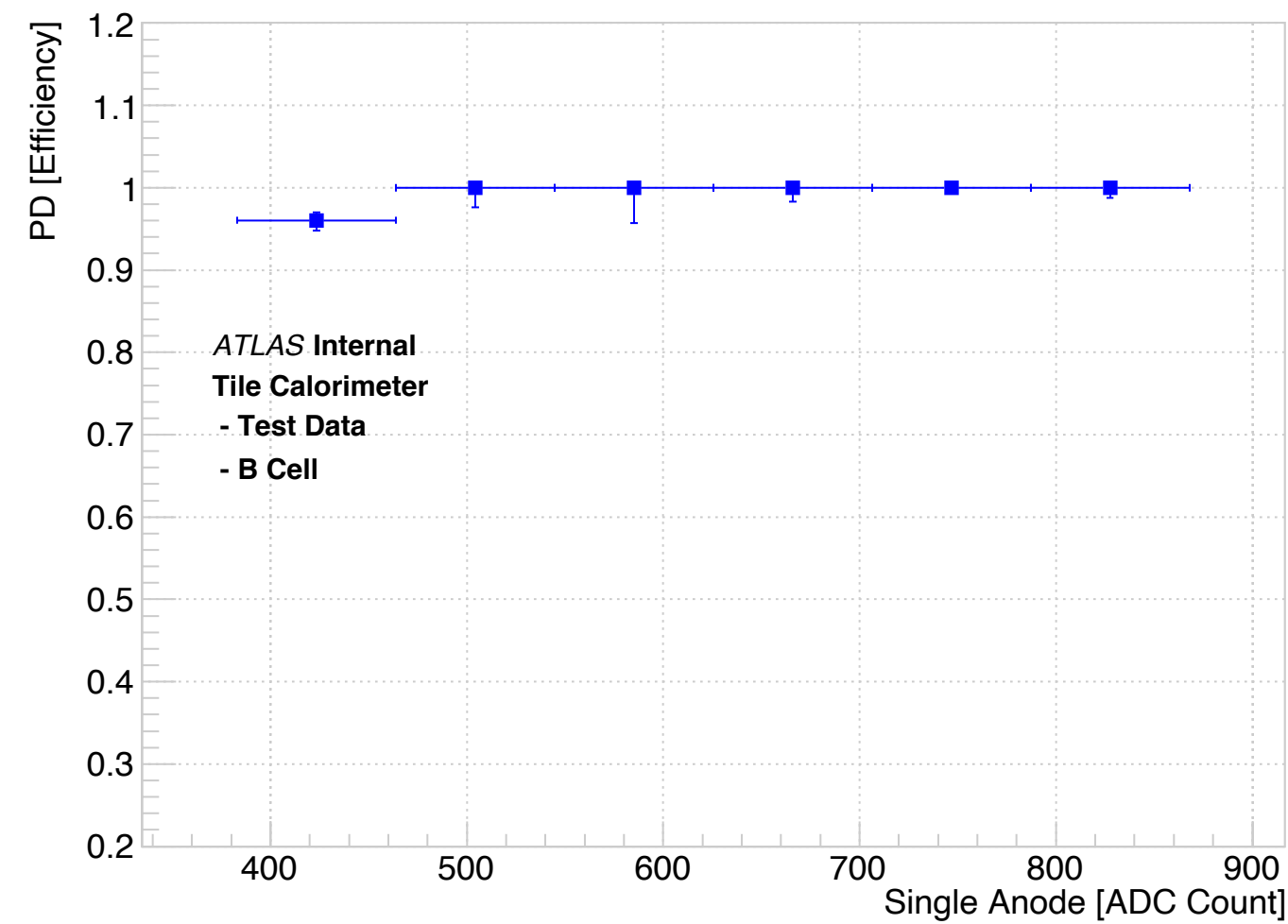


Fine Tunning



- Assim como na classificação da célula A2, o pipeline de *Fine Tuning* apresentou o melhor resultado entre os três pipelines propostos.

Classificação - Célula BC1



- Resultados na classificação da célula BC1 mostram eficiência máxima em todos os bins, com exceção de um único bin de baixa amplitude.
- A eficiência para a célula BC valida o processo de análise pois recupera a informação de células B e C separadas, assim como no projeto original do calorímetro.

Sumário - Resultados da Classificação

Resultados no Conjunto de Teste		
	Célula A2	Célula BC1
ACC	89.4 +- 3.3	97.1 +- 1.0
AUC	99.1 +- 0.3	99.9 +- 0.0
PD (Left/B)	98.3 +- 0.6	98.3 +- 0.9
FA (Left/B)	4.9 +- 1.1	1.4 +- 0.4
PD (Right/C)	95.1 +- 1.1	98.5 +- 0.4
FA (Right/C)	1.7 +- 0.6	1.7 +- 0.9

- Aqui mostramos as medidas de performance calculadas nas partições de teste (dados reais coletados durante a calibração de Cs).
- Granularidade longitudinal (células BC) possui resultados um pouco melhores que a granularidade transversal (células A).
- Isso acontece devido à separação natural das células BC em duas sub-células B e C separadamente.
- Medidas de Performance:
 - ACC - Acurácia
 - AUC - Área sob a curva ROC
 - PD - Probabilidade de detecção
 - FA - Falso Alarme

Sumário

- Usar uma combinação de modelos generativos e de classificação possibilitou separar células do TileCal em novas sub-células baseando-se no padrão de imagem coletado na grid de pixels da MA-PMT.
- Treinar modelos de CNN com uma combinação de diferentes pipelines de dados reais e sintéticos permitiu avaliar o impacto desses pipelines na incerteza de classificação.

Planos Futuros

- Mover a análise para dados coletados com o *test beam*.
- Analisar o impacto das imagens sintéticas na reconstrução de energia.
- Possibilidade de usar uma SiPMT no lugar da MA-PMT.

Obrigado
