

Análise Resultados

As análises apresentadas são relativas às duas observações principais feitas no documento:

1. O tempo que uma máquina leva para fazer o processamento da validação mais o tempo da primeira iteração da primeira época é bem menor que o tempo que se leva para executar o resto das iterações
2. O tempo que uma máquina leva para executar as suas iterações é bastante estável, tirando o tempo da primeira iteração da primeira época.

Os resultados apresentados são relativos aos experimentos:

- CNN / MNIST:
 - Rodado 3 vezes com 5 épocas.
 - Instâncias utilizadas em todas as vezes que foi executado: p2.xlarge (p2-1), p2.8xlarge (p2-8), p3.2xlarge (p3-1), p3.8xlarge (p3-4).
- FCN / Seismic:
 - Rodado 3 vezes com 5 épocas.
 - Instâncias utilizadas em todas as vezes que foi executado: p2.xlarge, p2.8xlarge, p2.16xlarge (p2-16), p3.2xlarge, p3.8xlarge, p3.16xlarge (p3-8).
- Resnet50 / Cifar10:
 - Rodado 3 vezes com 5 épocas.
 - Instâncias utilizadas em todas as vezes que foi executado: p2.xlarge, p2.8xlarge, p3.2xlarge, p3.8xlarge

Observação 1

Determinamos que o

$$Beta (\beta) = \frac{(tempo\ da\ primeira\ iteração + tempo\ de\ validação)}{(tempo\ das\ iterações\ remanescentes)}$$

mostra o quão significativos são os valores da primeira iteração e das validações são para o treinamento todo.

Também determinamos que o TMiAvg é o tempo médio das iterações remanescentes (tempo que das iterações sem contar a primeira).

Assim, com valores de Betas pequenos, temos que podemos usar o TMiAvg como métrica para comparar o tempo total de um treinamento entre instâncias diferentes.

1.1 Tabela Betas

Como podemos ver pelos dados abaixo, os valores dos Betas em todos os treinamentos foram elevados, isso devido ao fato que todos os treinamentos apresentados tiveram 5 épocas, e assim, o número de iterações não foi o bastante para que os valores da primeira iteração e das validações tivessem uma importância reduzida para o tempo de treinamento total.

Vemos uma tendência para Betas maiores para instâncias com um maior número de GPU's e em batch sizes maiores, esse último pode ser explicado pelo número de iterações, que é menor quando tempo batch sizes maiores.

1.1.1 CNN

	256		512		1024		2048	
Instância	Média (%)	Desvio padrão (%)	Média (%)	Desvio padrão (%)	Média (%)	Desvio padrão (%)	Média (%)	Desvio padrão (%)
p2-1	15.50%	1.68%	19.25%	2.76%	20.99%	0.16%	23.89%	0.54%
p2-8	74.54%	2.82%	114.15%	2.39%	162.08%	13.54%	207.52%	9.70%
p3-1	43.93%	3.18%	61.39%	3.88%	67.80%	4.18%	82.31%	4.34%
p3-4	103.42%	8.11%	190.14%	6.92%	273.46%	21.12%	376.22%	34.80%

1.1.2 FCN

	256		512		1024		2048	
Instância	Média (%)	Desvio padrão	Média (%)	Desvio padrão	Média (%)	Desvio padrão	Média (%)	Desvio padrão
p2-1	15.79%	13.22%	12.67%	3.77%	15.23%	1.77%	15.58%	0.97%
p2-8	27.63%	3.49%	47.32%	6.73%	56.75%	8.25%	56.81%	11.88%
p2-16	34.96%	0.33%	52.45%	8.36%	90.73%	0.96%	102.37%	14.78%
p3-1	26.73%	2.18%	29.70%	5.75%	30.00%	2.31%	31.92%	0.73%
p3-4	82.21%	2.47%	100.87%	9.26%	116.56%	10.04%	102.37%	19.17%
p3-8	83.05%	1.39%	124.52%	9.45%	178.91%	5.56%	179.13%	7.88%

1.1.3 ResNet50

	256		512		1024		2048	
Instância	Média (%)	Desvio padrão (%)	Média (%)	Desvio padrão (%)	Média (%)	Desvio padrão (%)	Média (%)	Desvio padrão (%)
p2-1	12.82%	0.34%	17.19%	0.48%	19.78%	0.34%	24.07%	0.40%
p2-8	96.88%	1.06%	162.18%	14.61%	230.18%	13.38%	312.92%	6.84%
p3-1	41.23%	1.17%	58.15%	1.29%	69.84%	4.19%	79.27%	1.61%
p3-4	141.03%	2.71%	241.29%	1.91%	386.65%	2.17%	570.40%	14.09%

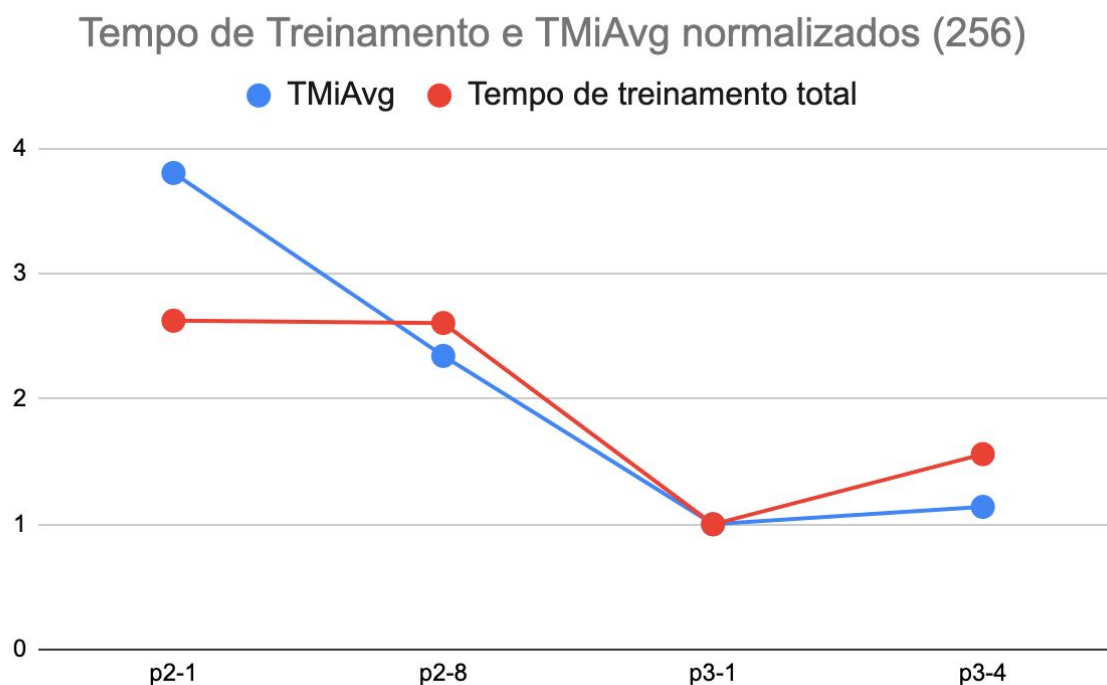
1.2 Comparativo entre a média do TMIAvg e do Tempo de treinamento (normalizados):

Normalizando o TMIAvg e o Tempo de treinamento (de acordo com o menor para cada batch) podemos ver se nesses treinamentos o TMIAvg se mostrou uma métrica válida da comparação dos tempos totais de treinamento entre as instâncias.

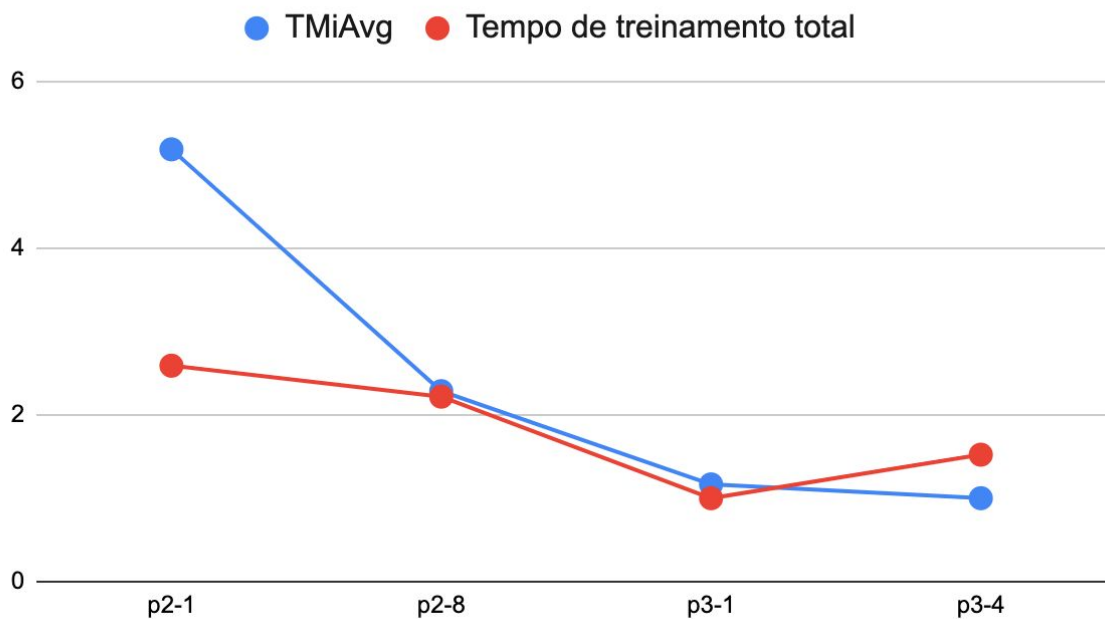
Vemos que não temos uma correlação precisa entre o TMIAvg e o tempo de treinamento total, mas no caso era o esperado para um número pequeno de épocas.

No experimento da CNN / MNist que roda por 30 épocas (1.2.4) vemos que temos uma correlação maior entre esses dois valores.

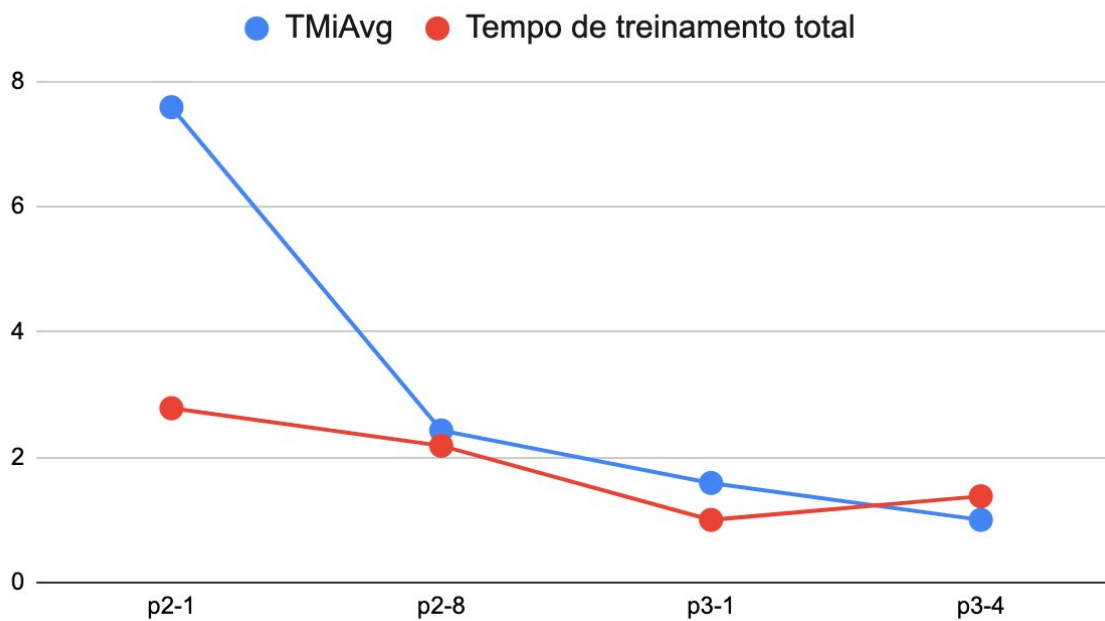
1.2.1 CNN



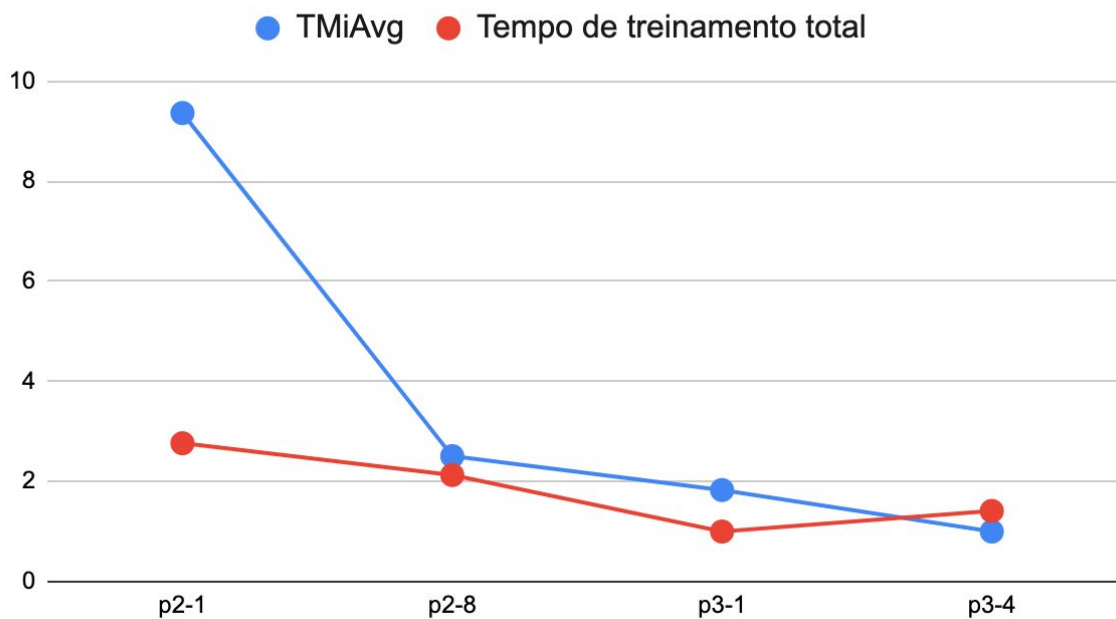
Tempo de Treinamento e TMiAvg normalizados (512)



Tempo de Treinamento e TMiAvg normalizados (1024)

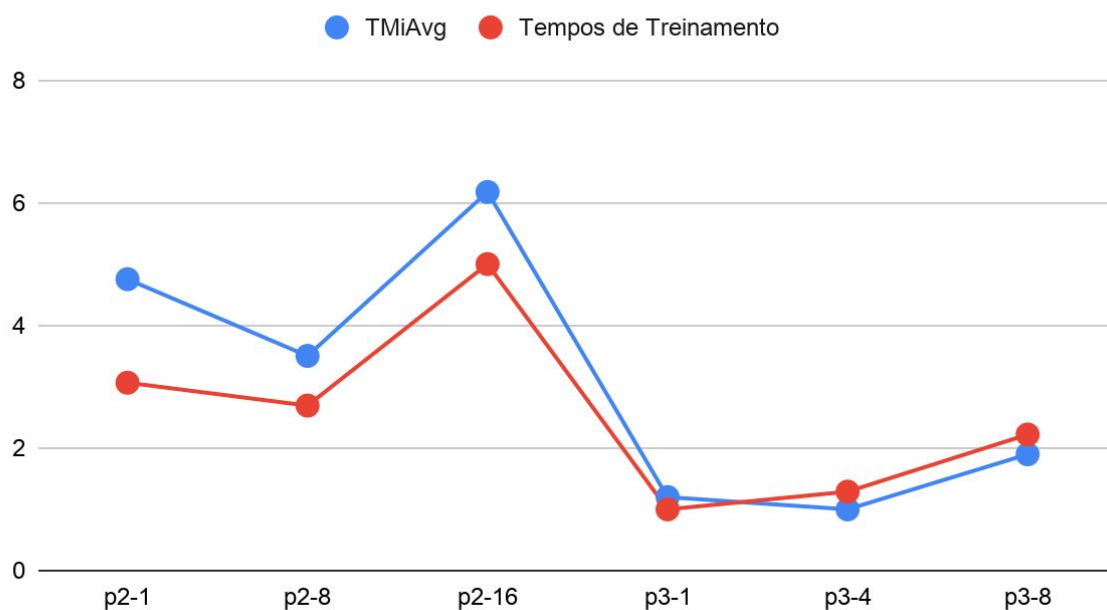


Tempo de Treinamento e TMiAvg normalizados (2048)

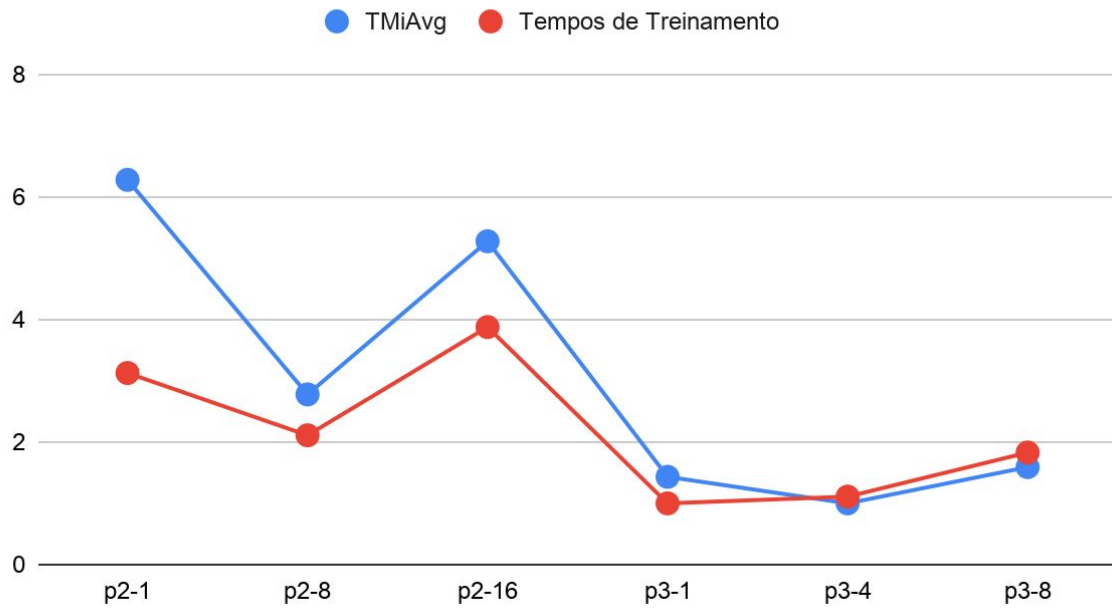


1.2.2 FCN

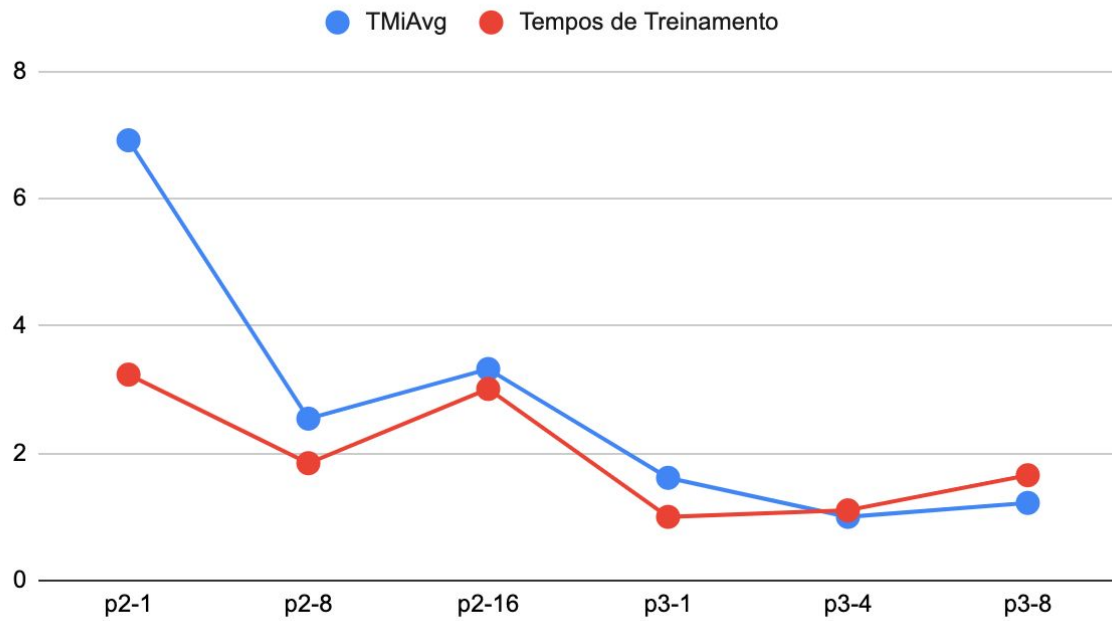
Tempos de Treinamento e TMiAvg normalizados (256)



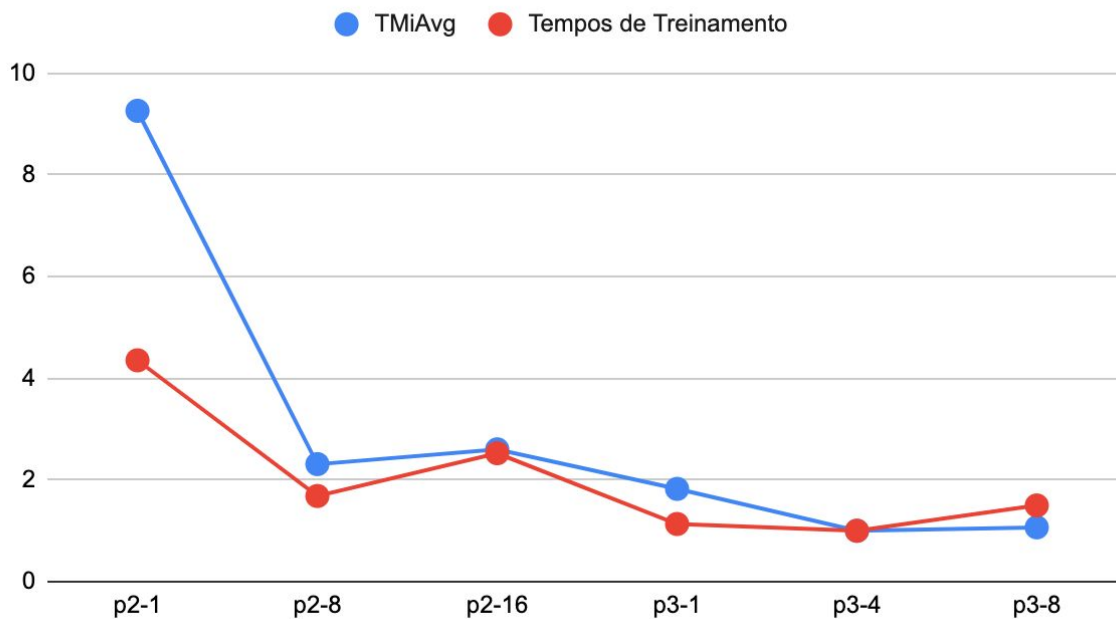
Tempos de Treinamento e TMiAvg normalizados (512)



Tempos de Treinamento e TMiAvg normalizados (1024)

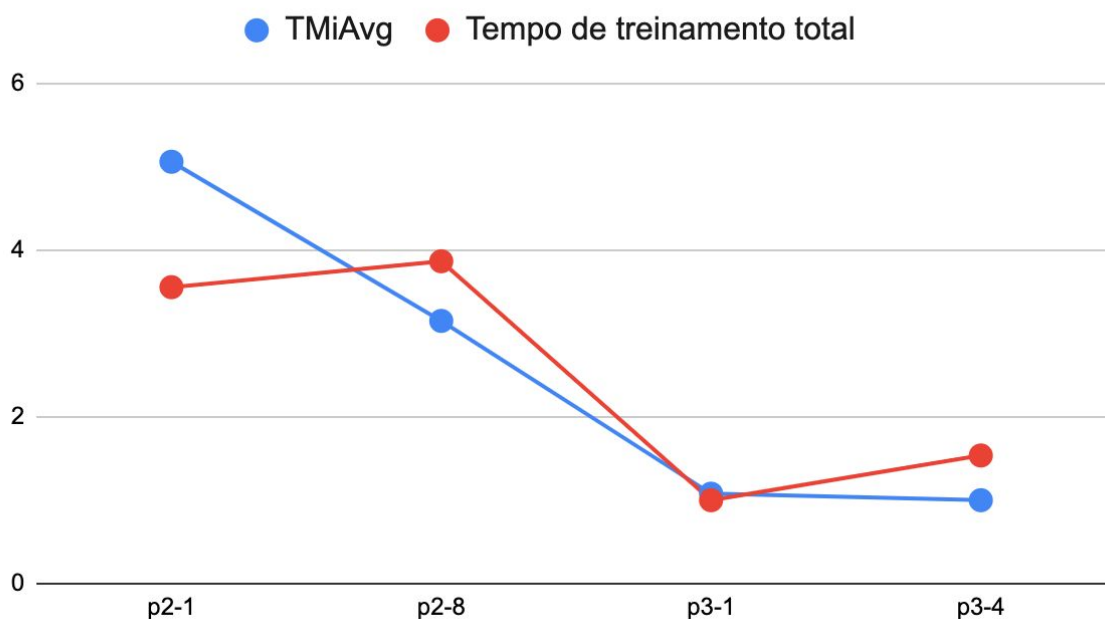


Tempos de Treinamento e TMiAvg normalizados (2048)

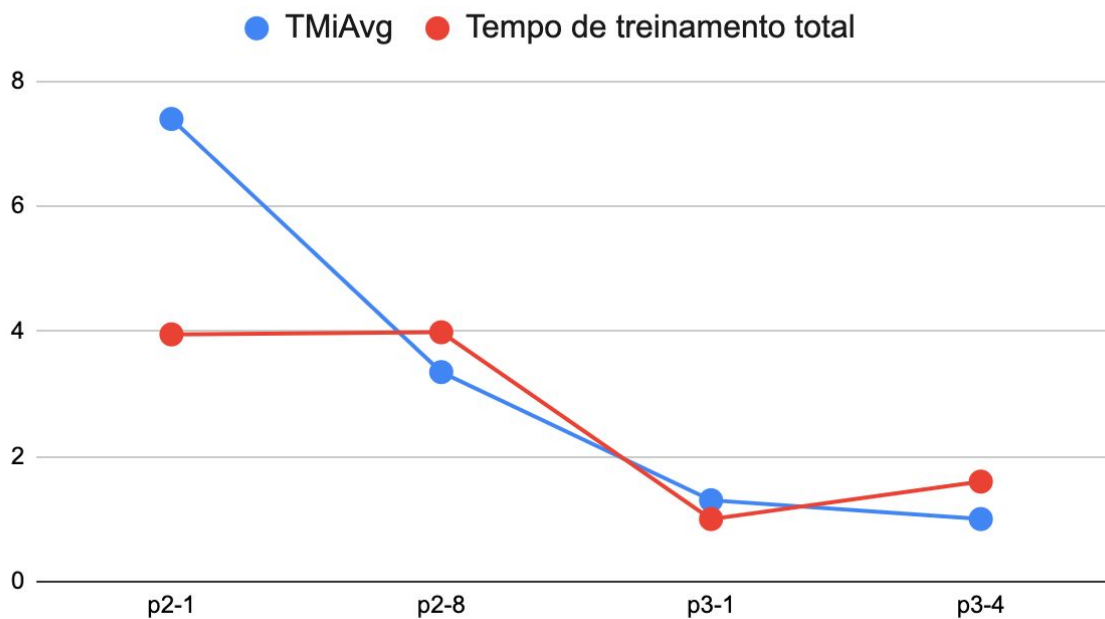


1.2.3 ResNet50

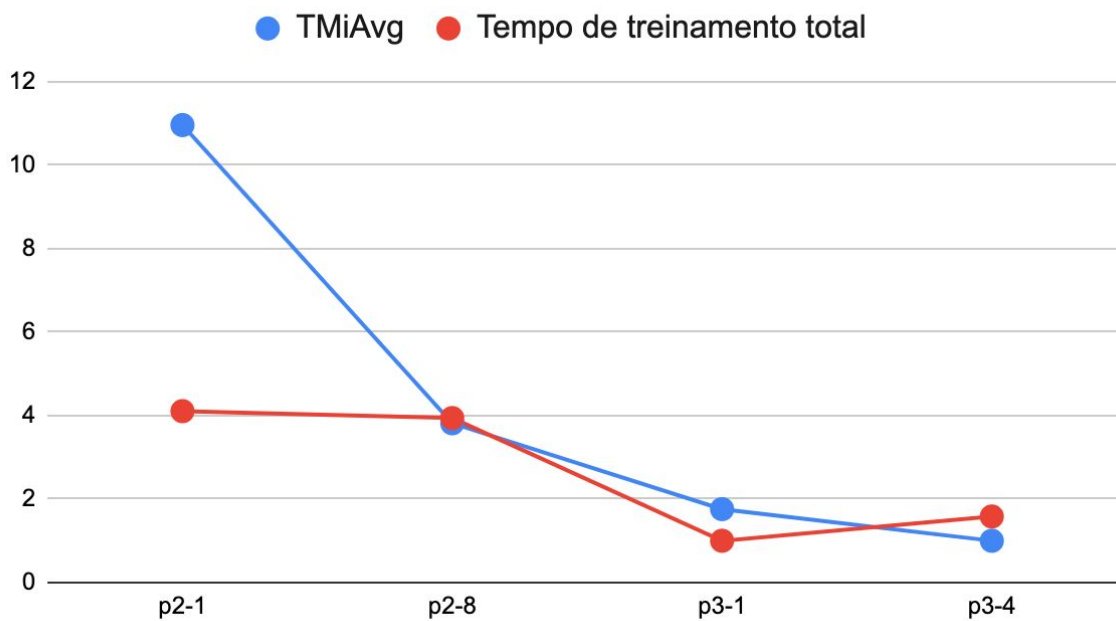
Tempo de Treinamento e TMiAvg normalizados (256)



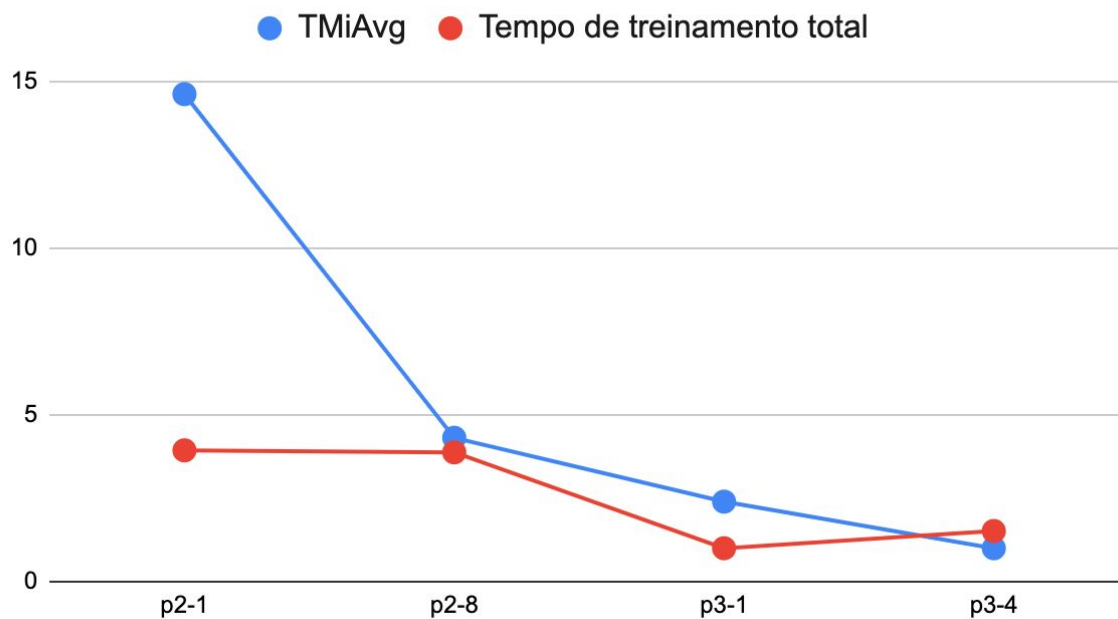
Tempo de Treinamento e TMiAvg normalizados (512)



Tempo de Treinamento e TMiAvg normalizados (1024)

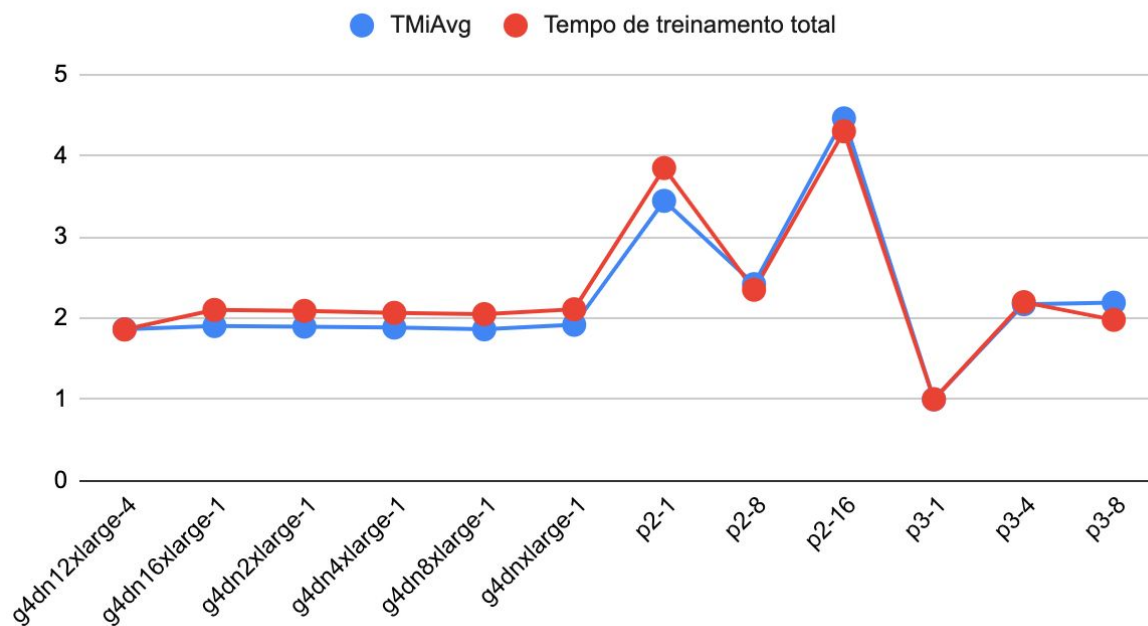


Tempo de Treinamento e TMiAvg normalizados (2048)

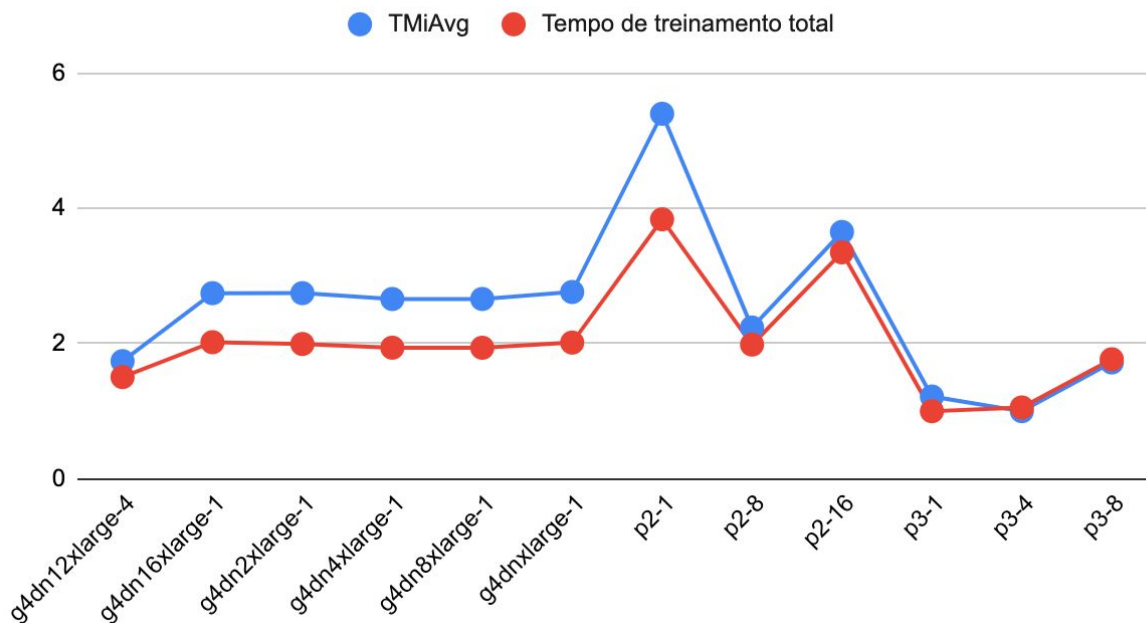


1.2.4 CNN - 30 epochs

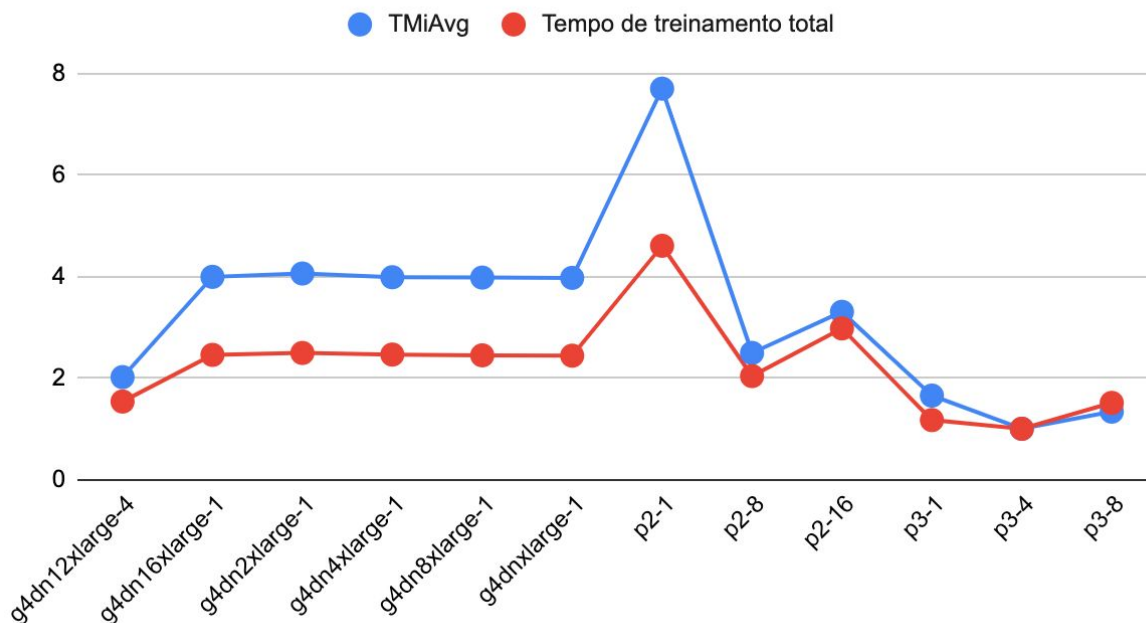
Tempo de Treinamento e TMiAvg normalizados (256)



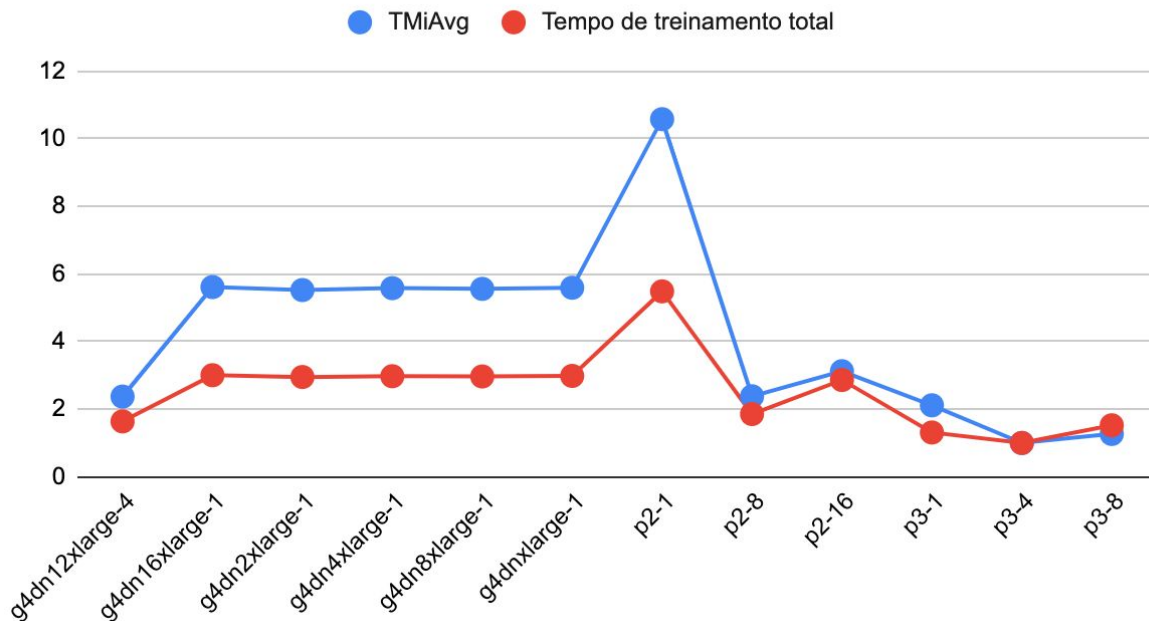
Tempo de Treinamento e TMiAvg normalizados (512)



Tempo de Treinamento e TMiAvg normalizados (1024)



Tempo de Treinamento e TMiAvg normalizados (2048)

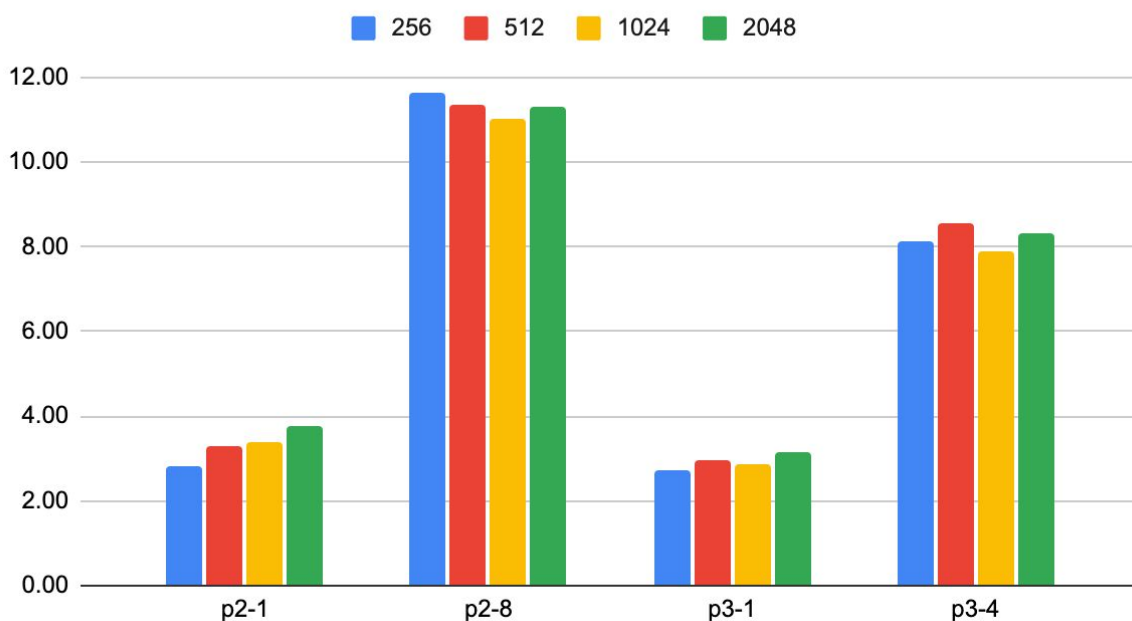


1.3 Relação primeira iteração - Número de GPU's

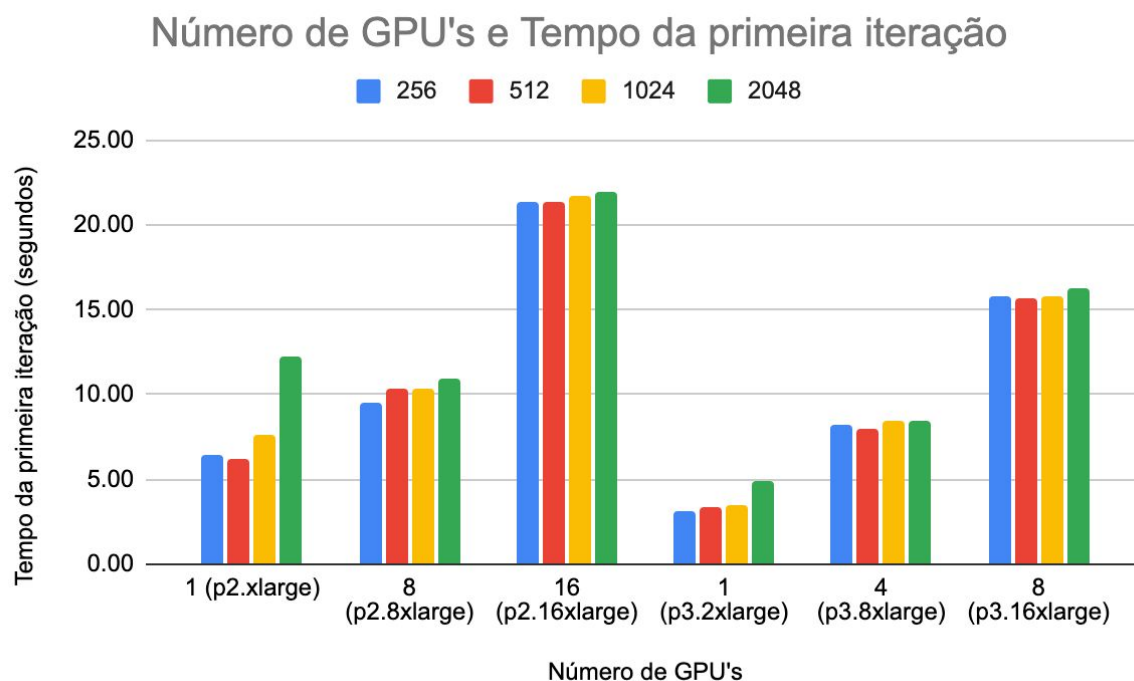
Outro fator que pode ser observado é a relação entre o número de GPU's numa instância e o tempo da primeira iteração do treinamento, nos gráficos abaixo temos o tempo médio da primeira iteração nos experimentos:

1.3.1 CNN

Número de GPU's e tempo da primeira iteração



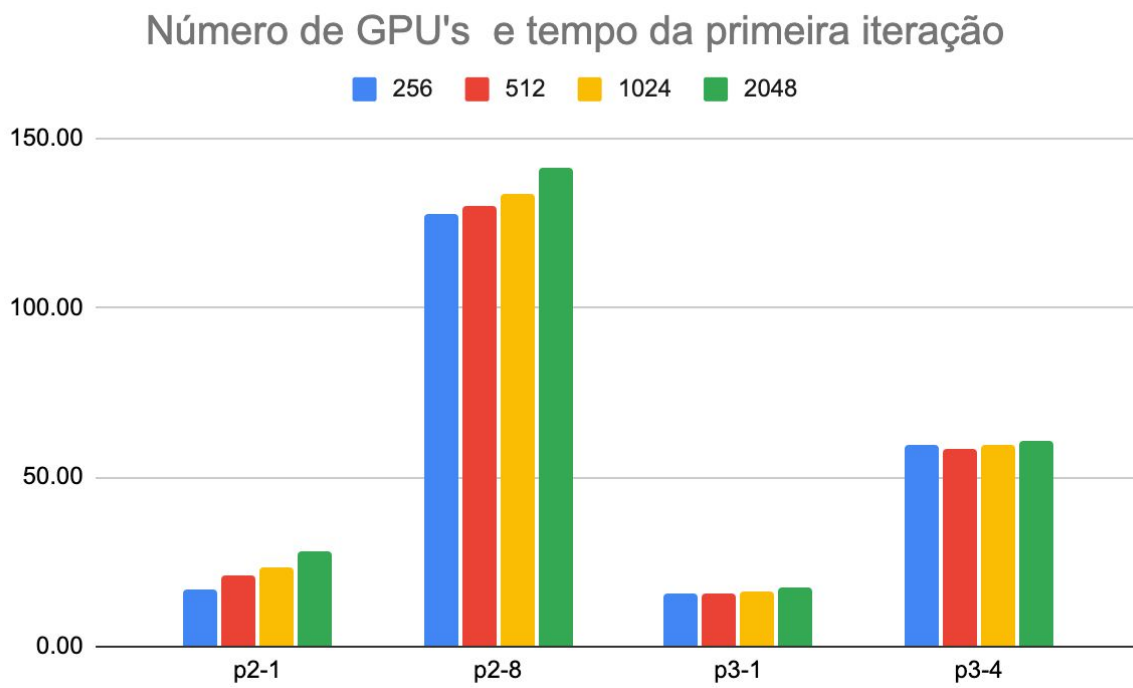
1.3.2 FCN



No caso dos experimentos com a FCN importante salientar que os tempos da primeira iteração na p2.xlarge tiveram um desvio padrão alto e que, no caso do batch size de 2048, ultrapassaram o tempo da p2.8xlarge:

Instância	256		512		1024		2048	
	Média (s)	Desvio padrão	Média (s)	Desvio padrão	Média (s)	Desvio padrão	Média (s)	Desvio padrão
p2-1	6.46	4.01	6.25	2.03	7.58	0.94	12.29	1.11
p2-8	9.53	1.40	10.31	1.73	10.30	1.32	10.89	1.61
p2-16	21.36	0.15	21.37	0.10	21.67	0.34	22.01	0.25
p3-1	3.16	0.20	3.35	0.74	3.47	0.18	4.93	0.26
p3-4	8.21	0.29	7.98	0.31	8.43	0.38	8.51	0.14
p3-8	15.77	0.14	15.73	0.09	15.85	0.18	16.22	0.55

1.3.3 ResNet50



Observação 2

O comportamento observado entre os experimentos relacionado à estabilidade do tempo das iterações (retirando) a primeira nos permite fazer estimativas para o TMIAvg com base nas primeiras iterações da primeira época, usamos três tipos de estimativa:

- Estimativa 1 (Step 2):
 - O tempo da segunda iteração usado para estimar o TMIAvg.
- Estimativa 2 (Avg 2-6):
 - O tempo médio entre a segunda e a sexta iteração é usado para estimar o TMIAvg.
- Estimativa 3 (Avg 2-11):
 - O tempo médio entre a segunda e a décima primeira iteração é usado para estimar o TMIAvg.

Para cada experimento, é mostrado a visão geral (média do erro entre as três vezes que foi executado de cada estimativa para cada batch size) e as tabelas de média e desvio padrão de cada estimativa.

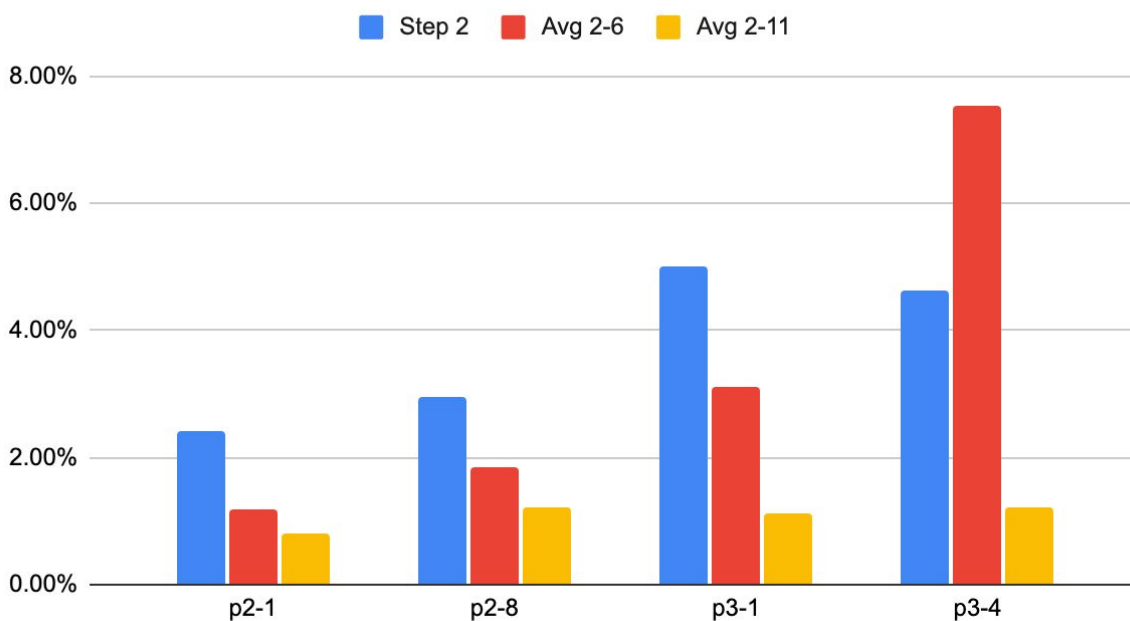
Pode ser observada uma tendência a erros maiores em batch sizes maiores pela Estimativa 3 (a mais precisa na maioria dos casos), isso porque, embora não seja contados o tempo da primeira iteração e os tempos das validações para o cálculo do TMIAvg, o tempo da última iteração da primeira época nos treinamentos é maior do que o das outras iterações, como é o caso das execuções da ResNet50 (2.1.3), que tivemos um erro médio alto em todas as estimativas, por motivos de um tempo alto dessa iteração em todas as suas execuções (cerca de 7,85 segundos em média, com um desvio padrão de 0,17). Como já dito antes, com um batch size maior temos menos iterações e assim esse tempo não é normalizado. Em experimentos como o com a FCN (2.1.2) com a instância p3.16xlarge (p3-8) com o batch size de 2048 vemos um erro médio com a Avg2-11 alto acompanhado de um alto desvio padrão, isso porque em duas das três execuções (e0 e e2) o tempo da primeira época (tirando a primeira iteração) foi mais rápido que das épocas subsequentes, na execução que isso não ocorreu (e1) tivemos um erro de 0.01%.

2.1 Estimativas

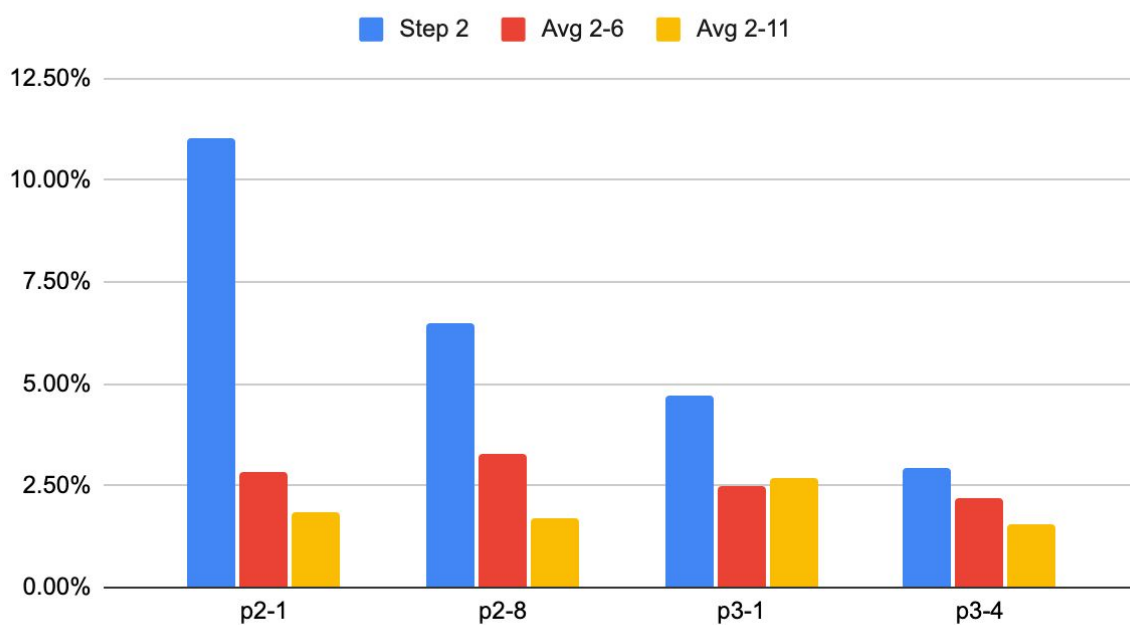
2.1.1 CNN

Visão Geral

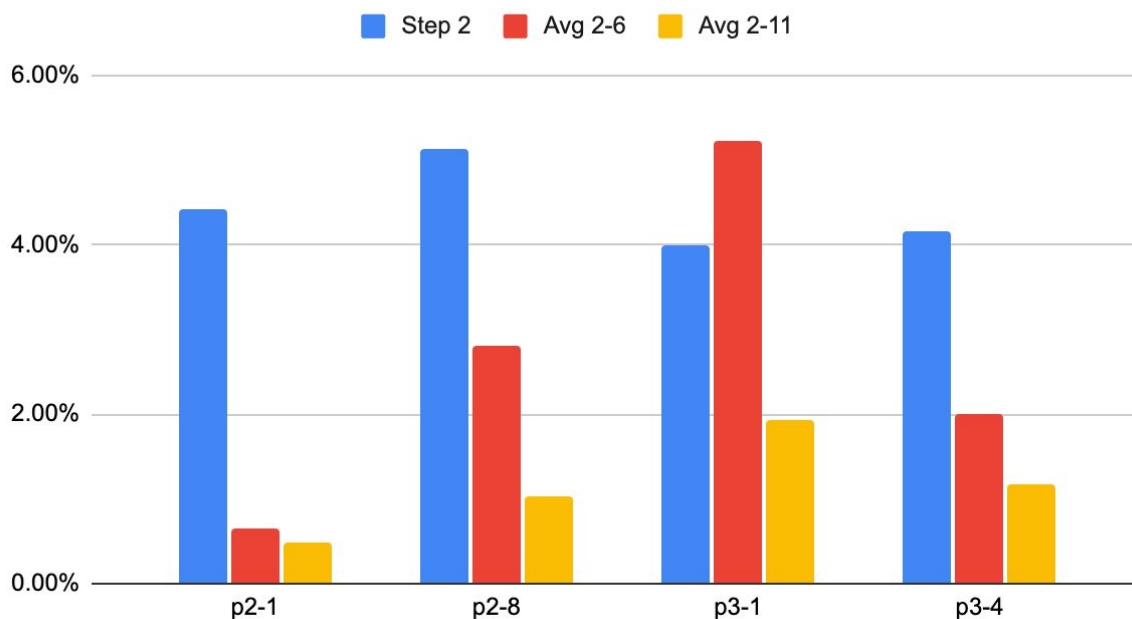
Média dos erros em relação ao TMiAvg (256)



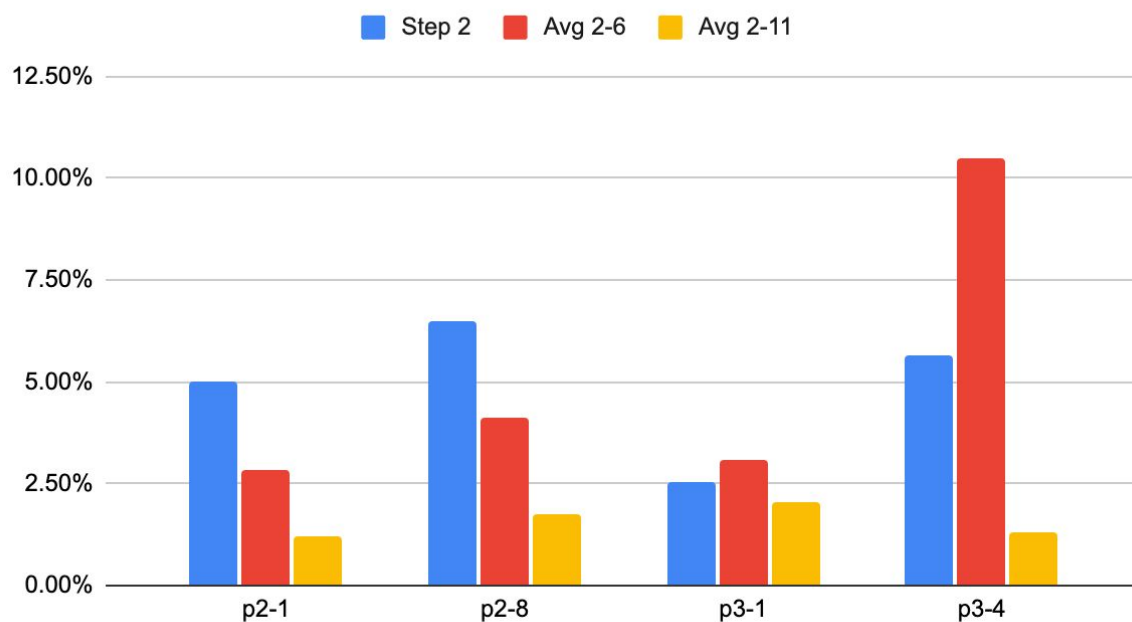
Média dos erros em relação ao TMiAvg (512)



Média dos erros em relação ao TMiAvg (1024)



Média dos erros em relação ao TMiAvg (2048)



Step 2

	256		512		1024		2048	
Instância	Média (%)	Desvio padrão	Média (s)	Desvio padrão	Média (s)	Desvio padrão	Média (s)	Desvio padrão
p2-1	2.42%	2.32%	11.03%	8.02%	4.43%	3.42%	5.03%	0.59%
p2-8	2.96%	4.14%	6.48%	0.00%	5.13%	4.47%	6.49%	5.31%
p3-1	5.01%	1.00%	4.70%	2.08%	4.00%	2.10%	2.56%	3.18%
p3-4	4.64%	4.25%	2.93%	0.23%	4.15%	4.64%	5.67%	4.55%

Avg 2-6

	256		512		1024		2048	
Instância	Média (%)	Desvio padrão	Média (s)	Desvio padrão	Média (s)	Desvio padrão	Média (s)	Desvio padrão
p2-1	1.19%	0.78%	2.86%	0.57%	0.64%	0.32%	2.83%	1.13%
p2-8	1.86%	2.55%	3.30%	0.00%	2.81%	3.41%	4.11%	4.53%
p3-1	3.12%	0.90%	2.47%	1.00%	5.23%	5.60%	3.08%	1.61%
p3-4	7.54%	5.08%	2.17%	1.31%	2.00%	1.24%	10.47%	7.93%

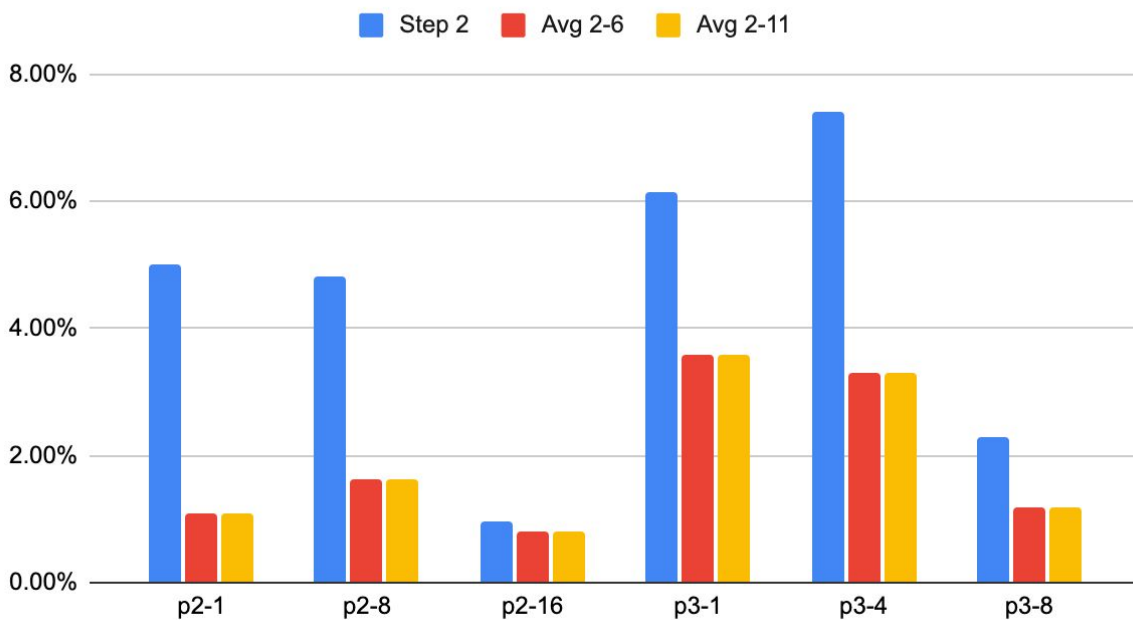
Avg 2-11

	256		512		1024		2048	
Instância	Média (%)	Desvio padrão	Média (s)	Desvio padrão	Média (s)	Desvio padrão	Média (s)	Desvio padrão
p2-1	0.80%	0.22%	1.86%	0.69%	0.50%	0.17%	1.19%	0.02%
p2-8	1.21%	0.36%	1.72%	0.42%	1.04%	1.31%	1.77%	1.02%
p3-1	1.11%	0.16%	2.67%	3.22%	1.94%	1.04%	2.03%	2.47%
p3-4	1.23%	0.77%	1.54%	1.97%	1.18%	0.40%	1.29%	0.85%

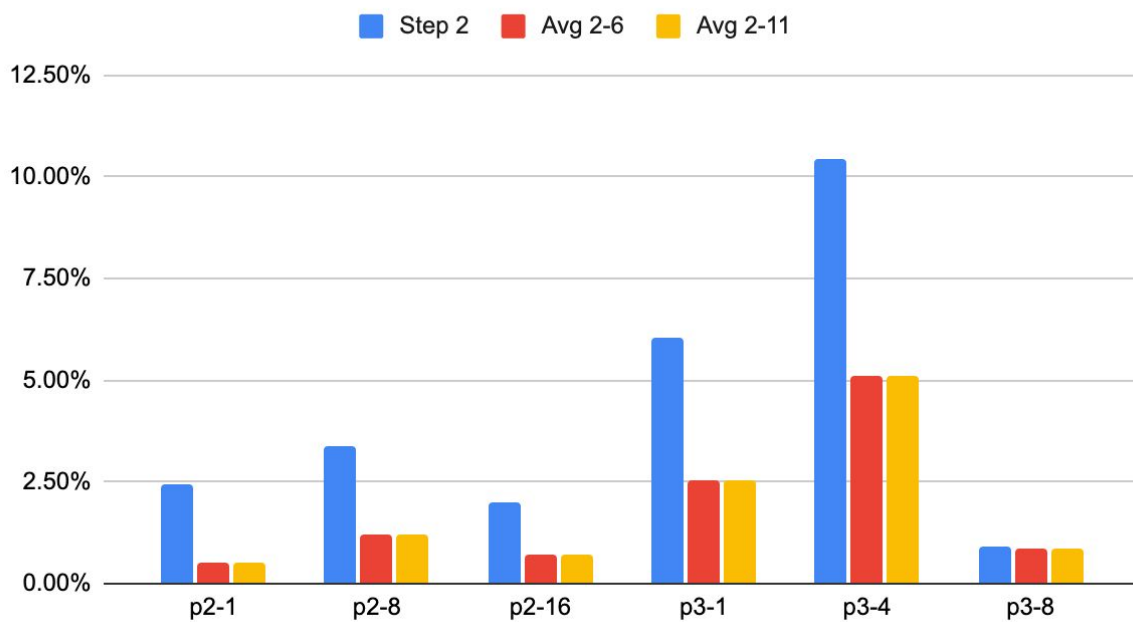
2.1.2 FCN

Visão Geral

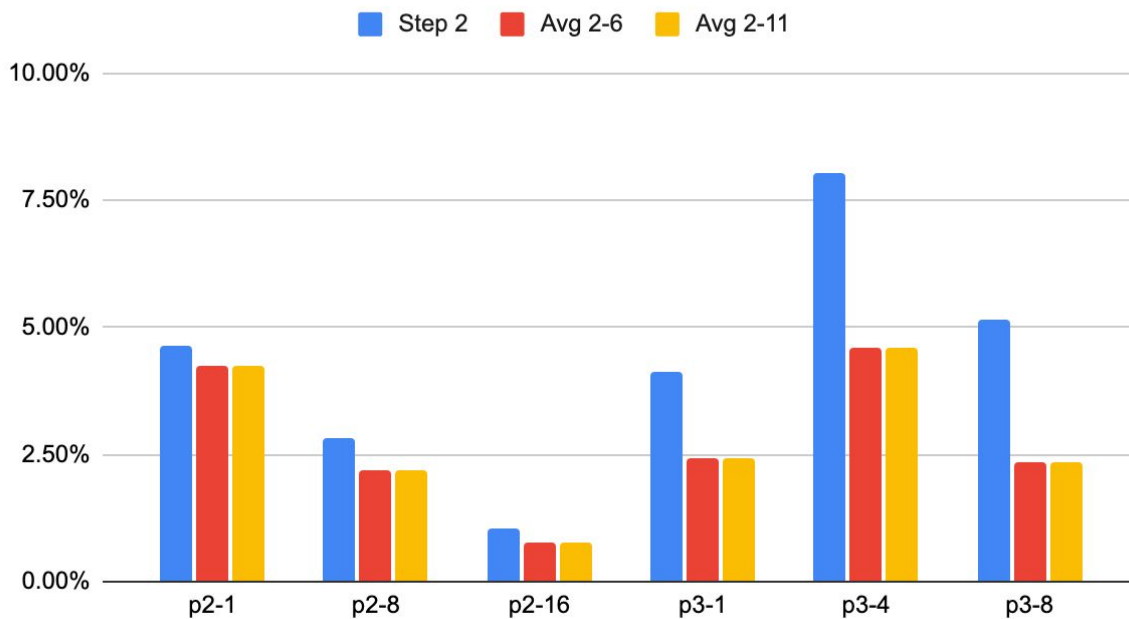
Média dos erros em relação ao TMiAvg (256)



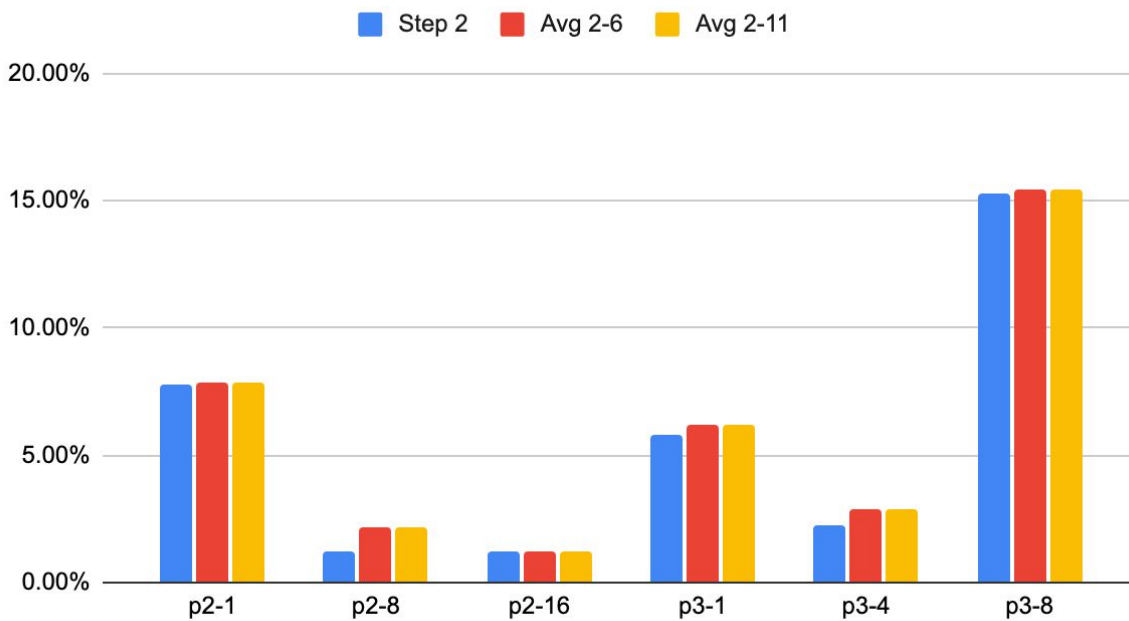
Média dos erros em relação ao TMiAvg (512)



Média dos erros em relação ao TMiAvg (1024)



Média dos erros em relação ao TMiAvg (2048)



Step 2

	256		512		1024		2048	
Instância	Média (%)	Desvio padrão	Média (%)	Desvio padrão	Média (%)	Desvio padrão	Média (%)	Desvio padrão
p2-1	5.02%	3.42%	2.44%	2.95%	4.63%	1.28%	7.74%	1.00%
p2-8	4.81%	6.12%	3.37%	4.93%	2.81%	0.83%	1.24%	0.80%
p2-16	0.96%	1.25%	2.02%	2.46%	1.05%	0.76%	1.21%	1.01%
p3-1	6.14%	4.61%	6.04%	6.40%	4.13%	1.18%	5.83%	2.18%
p3-4	7.42%	3.57%	10.43%	10.09%	8.02%	6.35%	2.28%	1.27%
p3-8	2.30%	2.01%	0.92%	0.74%	5.17%	5.37%	15.30%	12.60%

Avg 2-6

	256		512		1024		2048	
Instância	Média (%)	Desvio padrão	Média (%)	Desvio padrão	Média (%)	Desvio padrão	Média (%)	Desvio padrão
p2-1	1.08%	0.74%	0.51%	0.31%	4.23%	0.56%	7.87%	0.70%
p2-8	1.61%	1.04%	1.19%	0.45%	2.20%	2.65%	2.20%	1.25%
p2-16	0.80%	0.42%	0.74%	0.60%	0.76%	0.16%	1.24%	0.76%
p3-1	3.58%	2.14%	2.55%	1.11%	2.45%	0.32%	6.17%	1.89%
p3-4	3.31%	2.53%	5.09%	3.70%	4.59%	2.11%	2.85%	1.71%
p3-8	1.19%	0.87%	0.87%	0.59%	2.36%	1.08%	15.46%	12.43%

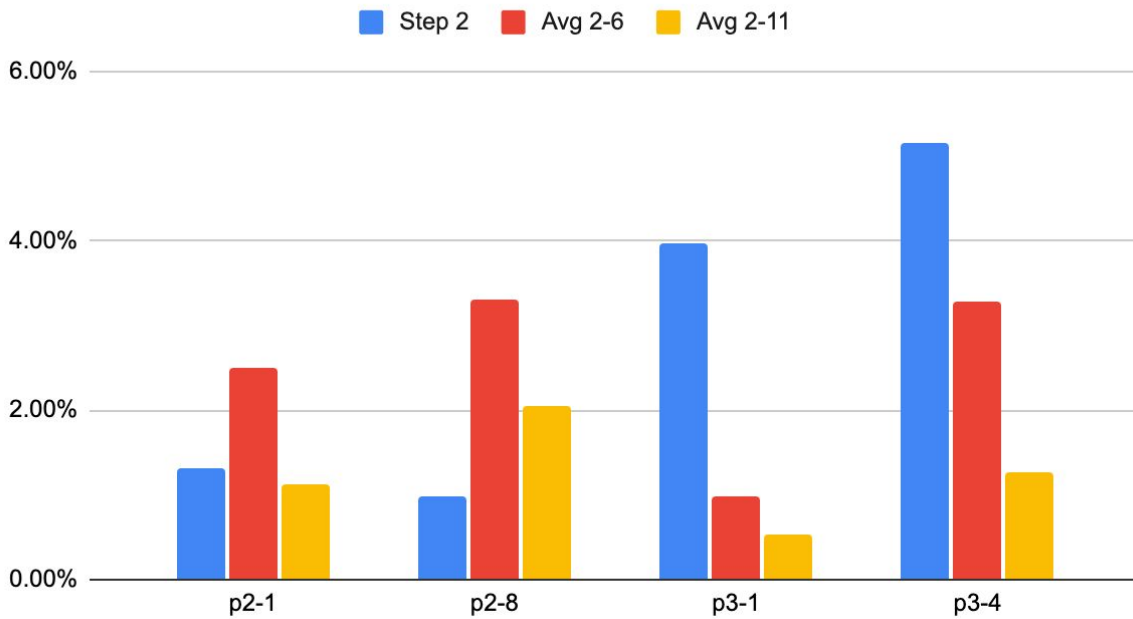
Avg 2-11

	256		512		1024		2048	
Instância	Média (%)	Desvio padrão	Média (%)	Desvio padrão	Média (%)	Desvio padrão	Média (%)	Desvio padrão
p2-1	0.26%	0.20%	0.28%	0.19%	4.65%	0.83%	7.81%	0.79%
p2-8	0.72%	0.79%	1.31%	0.81%	2.13%	3.06%	1.86%	0.77%
p2-16	0.61%	0.53%	0.39%	0.27%	1.08%	0.73%	1.04%	0.72%
p3-1	2.57%	1.51%	2.08%	1.35%	2.98%	0.09%	6.05%	1.73%
p3-4	3.40%	1.02%	5.38%	2.46%	3.62%	0.24%	2.43%	1.90%
p3-8	1.33%	0.88%	1.53%	0.57%	2.14%	0.45%	14.91%	12.91%

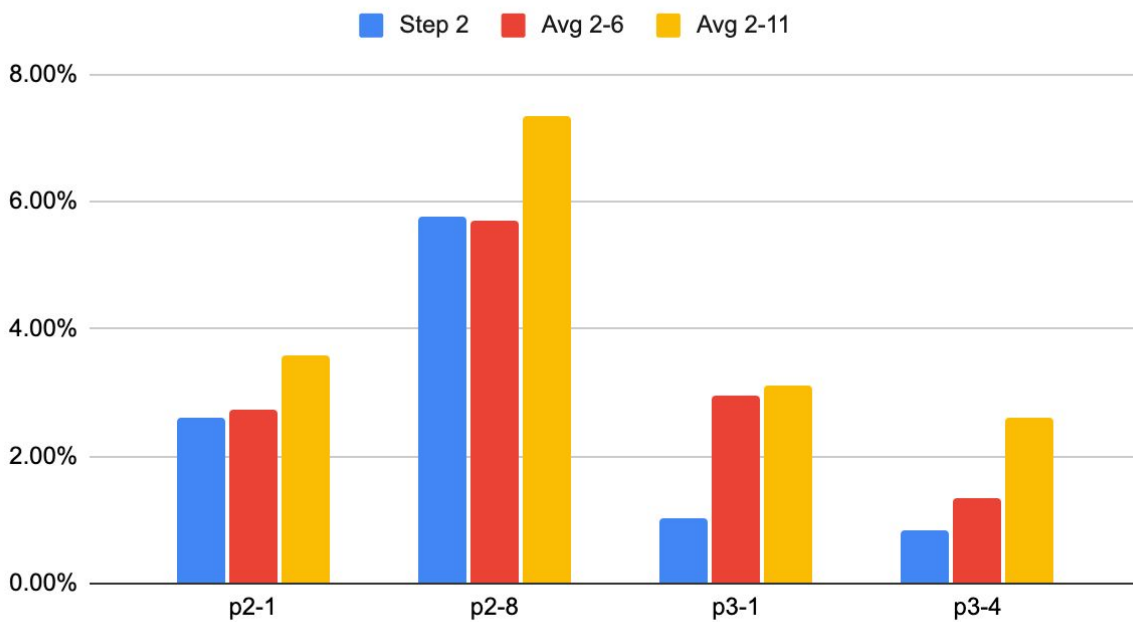
2.1.3 ResNet50

Visão Geral

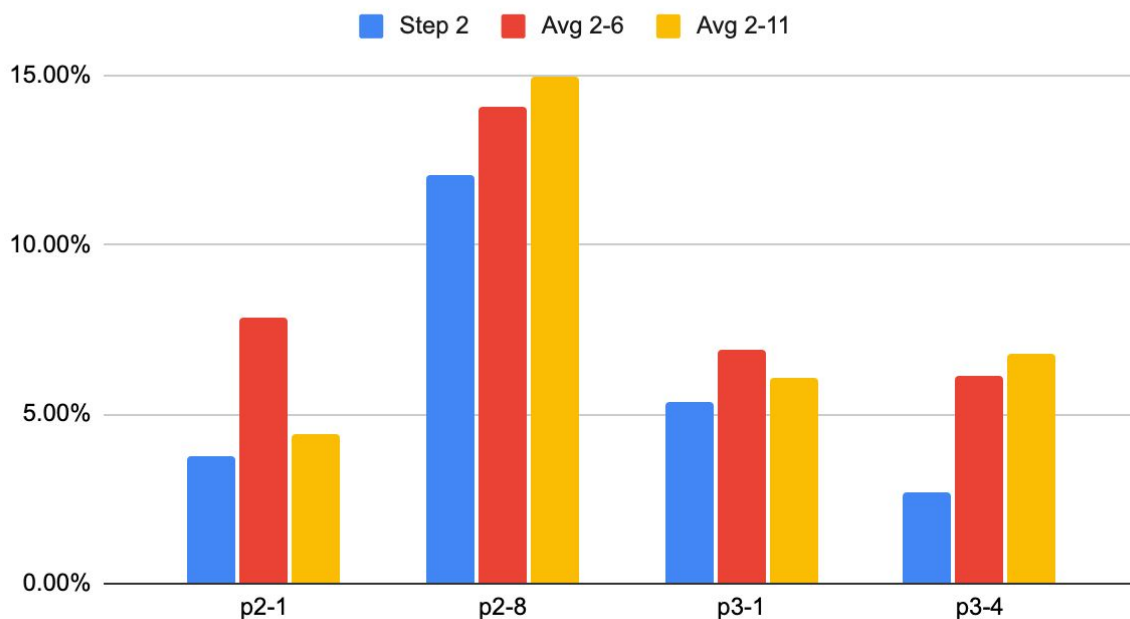
Média dos erros em relação ao TMiAvg (256)



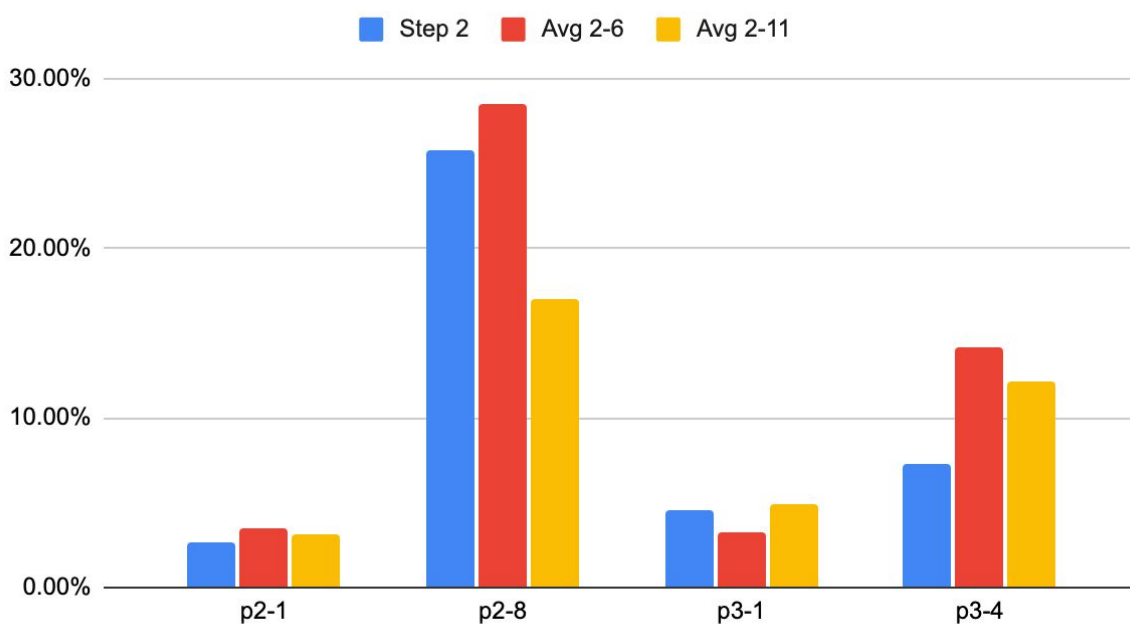
Média dos erros em relação ao TMiAvg (512)



Média dos erros em relação ao TMiAvg (1024)



Média dos erros em relação ao TMiAvg (2048)



Step 2

	256		512		1024		2048	
Instância	Média (%)	Desvio padrão	Média (s)	Desvio padrão	Média (s)	Desvio padrão	Média (s)	Desvio padrão
p2-1	1.33%	1.24%	2.59%	0.09%	3.74%	2.29%	2.70%	0.46%
p2-8	0.99%	0.83%	5.78%	3.11%	12.05%	9.65%	25.72%	11.31%
p3-1	3.96%	0.38%	1.01%	0.48%	5.35%	0.23%	4.59%	0.78%
p3-4	5.15%	1.54%	0.83%	0.60%	2.72%	0.99%	7.29%	2.76%

Avg 2-6

	256		512		1024		2048	
Instância	Média (%)	Desvio padrão	Média (s)	Desvio padrão	Média (s)	Desvio padrão	Média (s)	Desvio padrão
p2-1	2.50%	1.85%	2.72%	2.07%	7.86%	1.39%	3.51%	2.46%
p2-8	3.31%	2.28%	5.71%	0.25%	14.10%	2.70%	28.44%	0.47%
p3-1	0.99%	0.35%	2.95%	0.37%	6.89%	1.56%	3.31%	0.92%
p3-4	3.29%	2.84%	1.35%	1.42%	6.11%	0.92%	14.12%	3.93%

Avg 2-11

	256		512		1024		2048	
Instância	Média (%)	Desvio padrão	Média (s)	Desvio padrão	Média (s)	Desvio padrão	Média (s)	Desvio padrão
p2-1	1.12%	0.48%	3.58%	0.60%	4.40%	2.48%	3.10%	0.53%
p2-8	2.04%	0.31%	7.33%	3.15%	14.93%	1.16%	16.98%	3.89%
p3-1	0.53%	0.36%	3.11%	0.30%	6.09%	0.19%	4.89%	0.56%
p3-4	1.26%	0.79%	2.59%	0.52%	6.80%	1.15%	12.13%	0.18%