

FixMatch no CIFAR-10: implementação, experimentos e análise

Anderson Falcão

Leonardo Alexandre

Ramyro Corrêa

Introdução

- FixMatch reduz a necessidade de rótulos no CIFAR-10.
- Com poucos rótulos por classe, é possível alcançar boas acurácia alinhando pseudo-rótulos confiáveis com fortes aumentações.
- Pontos-chave: algoritmo semi-supervisionado (supervisionado + consistência), pseudo-rótulos só entram quando a confiança $\geq \tau$.
- Avaliamos 1, 4, 25, 70 e 400 rótulos/classe.

FixMatch

- x (não rotulado) \rightarrow weak aug $\rightarrow p(y|x^{\text{weak}})$
- Se $\max p \geq \tau \Rightarrow$ pseudo-rótulo $\hat{y} = \operatorname{argmax} p$
- Mesma x com strong aug \rightarrow treinar para prever \hat{y}
- Em paralelo: treino supervisionado nas amostras rotuladas
- Intuição: A versão fraca fornece um rótulo confiável; a versão forte força robustez.

LOSS Híbrida

$$\text{Loss} = \text{Loss_supervisionada} + \lambda_u \cdot \mathbf{1}[\text{confiança} \geq \tau] \cdot \text{Loss_não-supervisionada}$$

- Supervisionada: CE(rótulo real, predição)
- Não-supervisionada: CE(pseudo-rótulo, predição em strong aug)
- Hiperparâmetros-chave: τ (threshold), λ_u (peso), μ (ratio unlabeled:labeled por batch)

Ideia: unir um “sinal limpo” (rótulos verdadeiros) com um “sinal barato” (pseudo-rótulos confiáveis) para empurrar o modelo a ser consistente mesmo sob aumentações fortes, de modo que rótulos reais ancam o modelo; pseudo-rótulos espalham esse aprendizado para o mar de dados não rotulados, desde que sejam confiáveis.

Implementação

- Modelo base: ResNet-18
- Dados: CIFAR-10 (divisão em rotulados vs. não rotulados)
- Augmentações: flip/crop padrão para weak e RandAugment para strong.
- Treino: SGD+Momentum; $\mu \approx 7$
- Logging: acurácia, losses, acceptance rate de pseudo-rótulos

Experimentos

- Cenários: 1, 4, 25, 70 e 400 rótulos por classe (total: 10, 40, 250, 700, 4.000 dados rotulados no total), em um universo de 10.000 dados.
- Amostragem: estratificada por classe (mesma quantidade por classe)
- Configuração: $\mu \approx 7$, $\tau = 0,95$, $\lambda_u = 1$
- Métricas: acurácia de teste (%), loss de teste e acceptance rate (fração de pseudo-rótulos aceitos).

Resultados

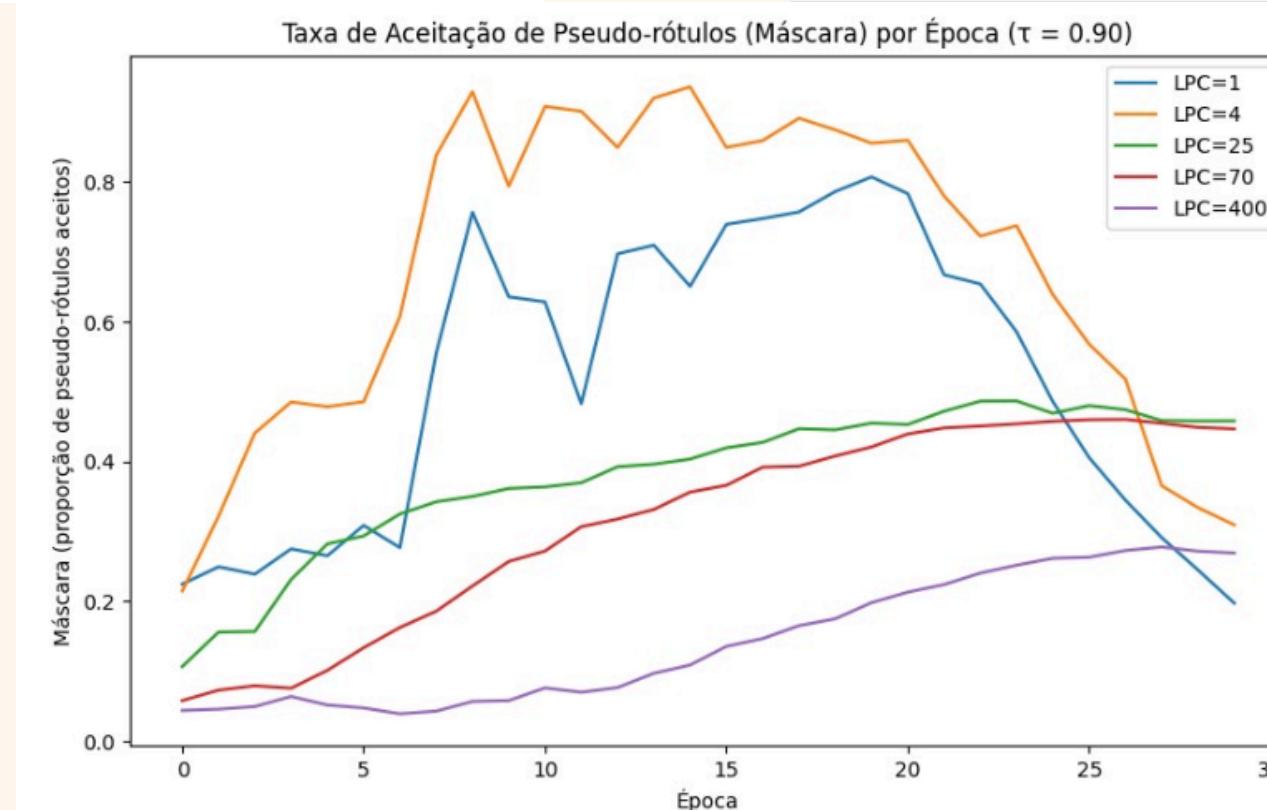
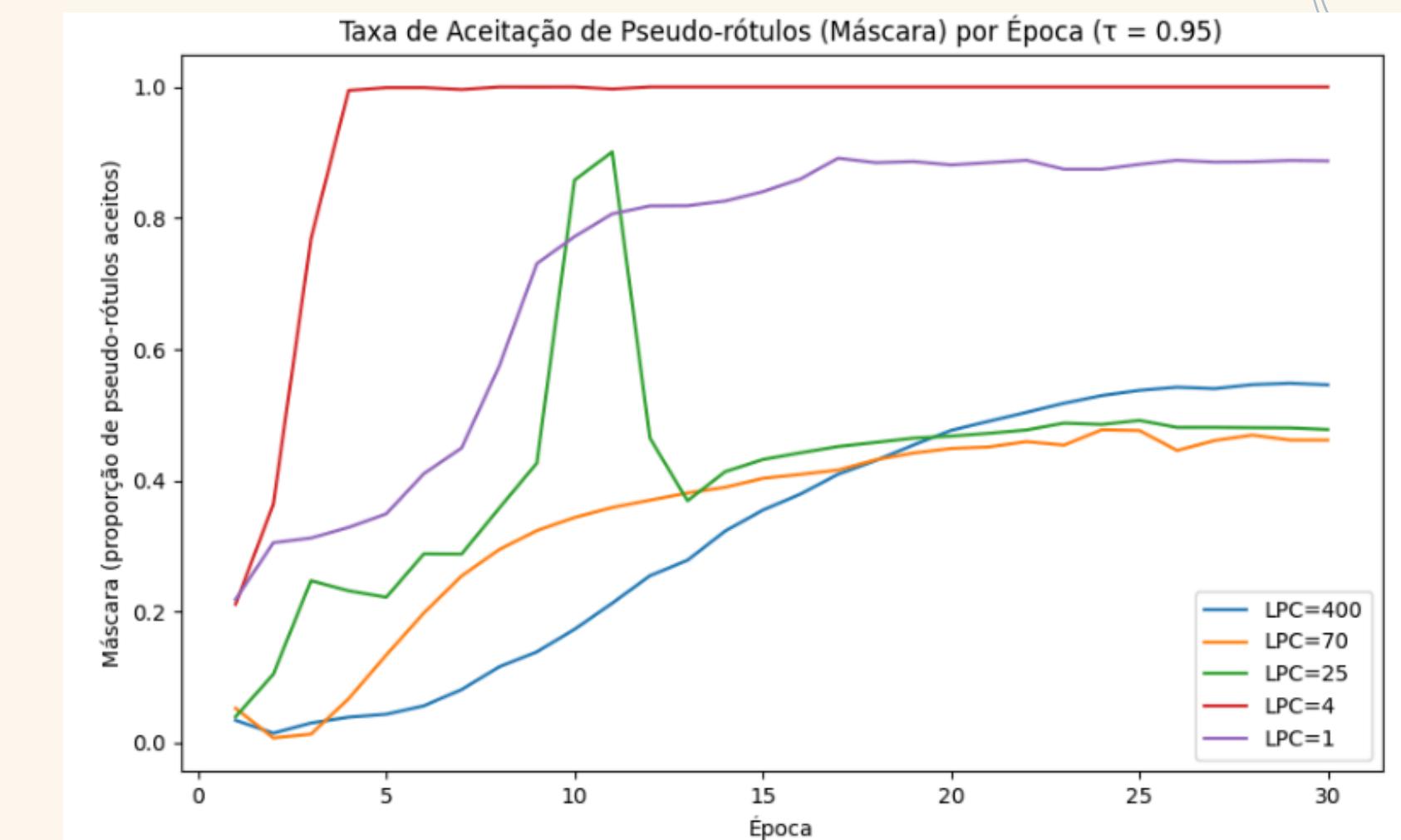
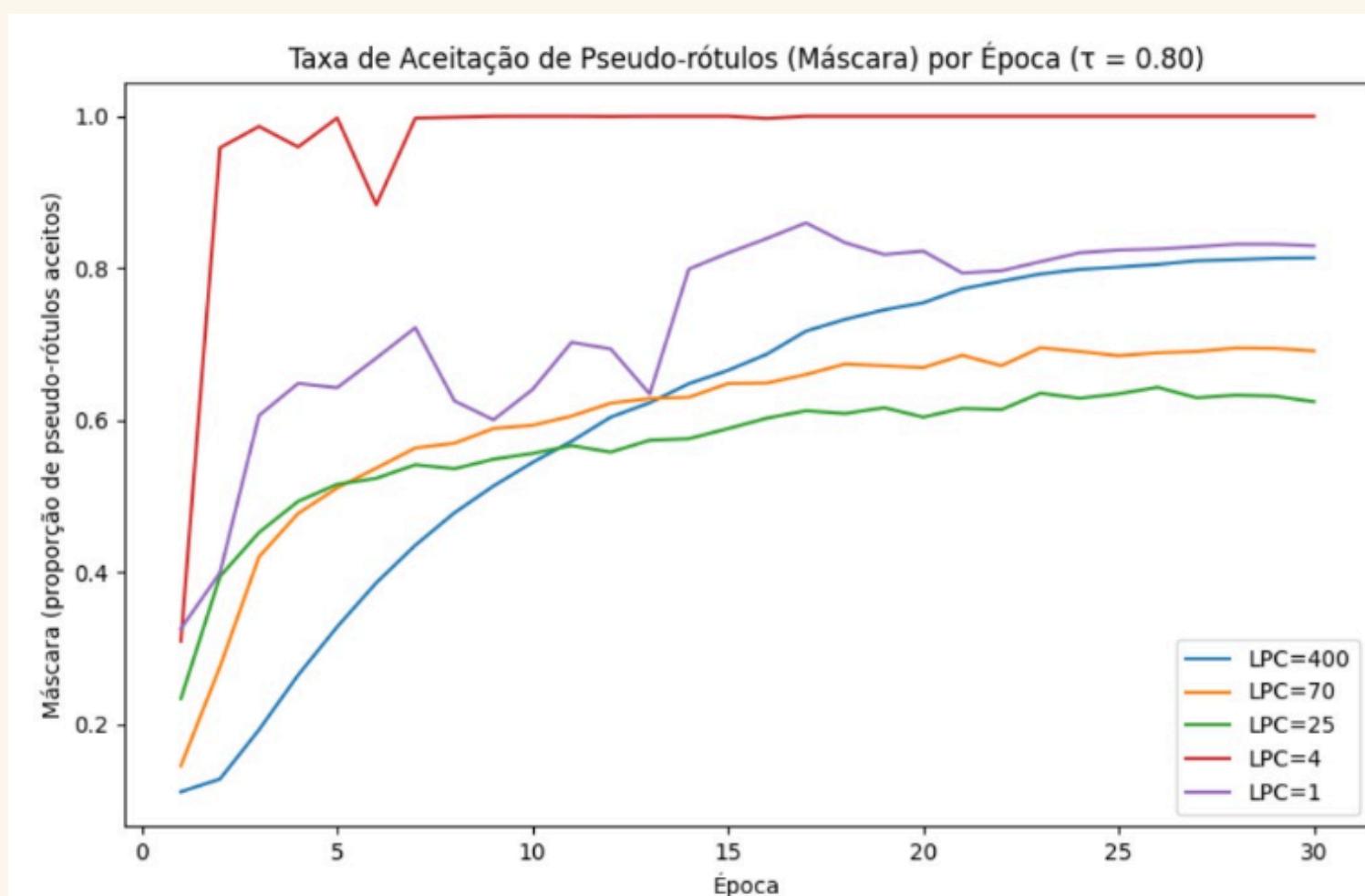
[Esperado] Com menos rótulos a curva é mais ruidosa e lenta, enquanto o acceptance rate tende a subir ao longo das épocas conforme o modelo melhora.

[Esperado] A qualidade/aceite por classe deve se tornar menos concentrada ao longo do treino, reduzindo colapso e equilibrando a cobertura entre classes.

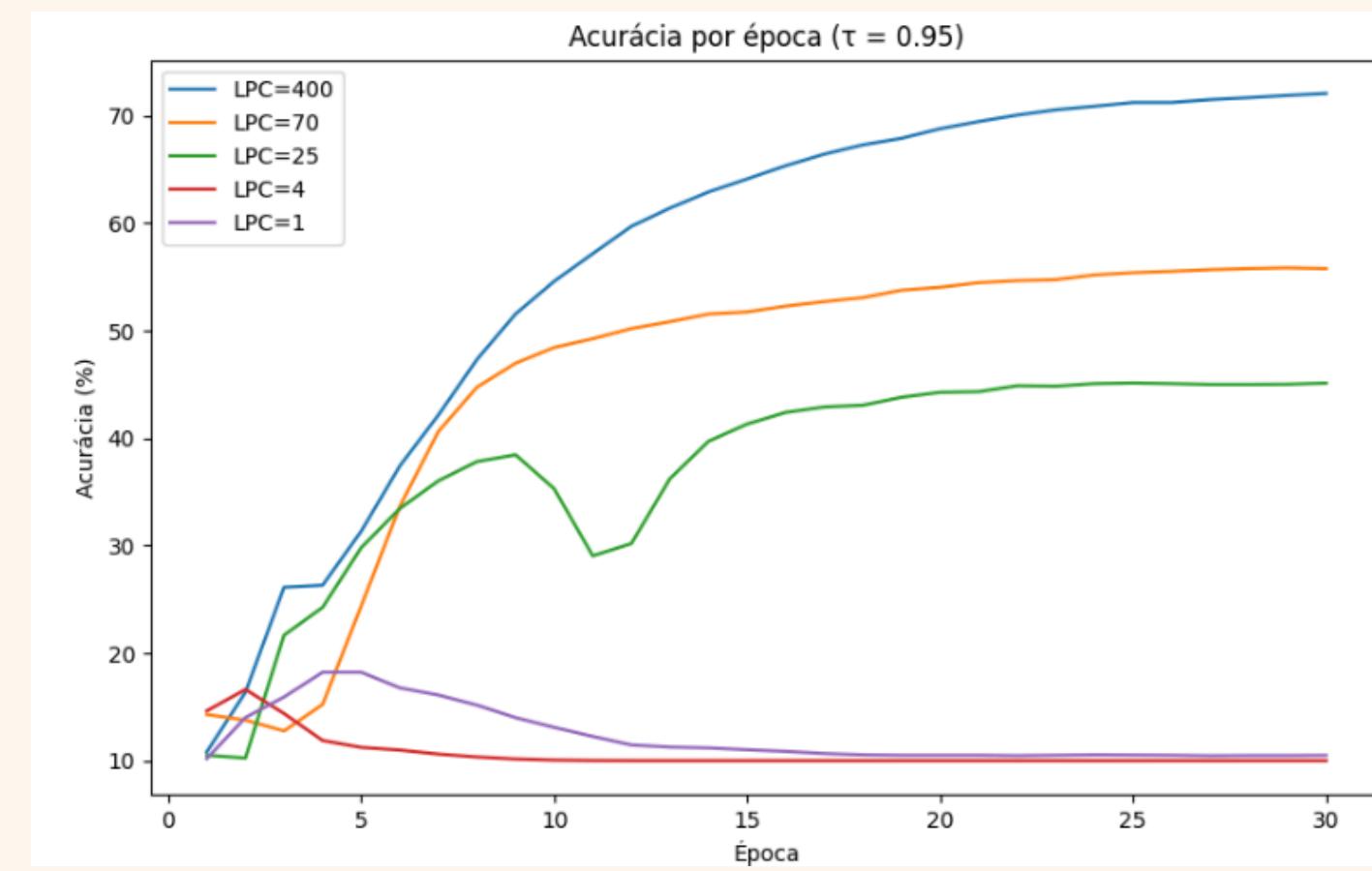
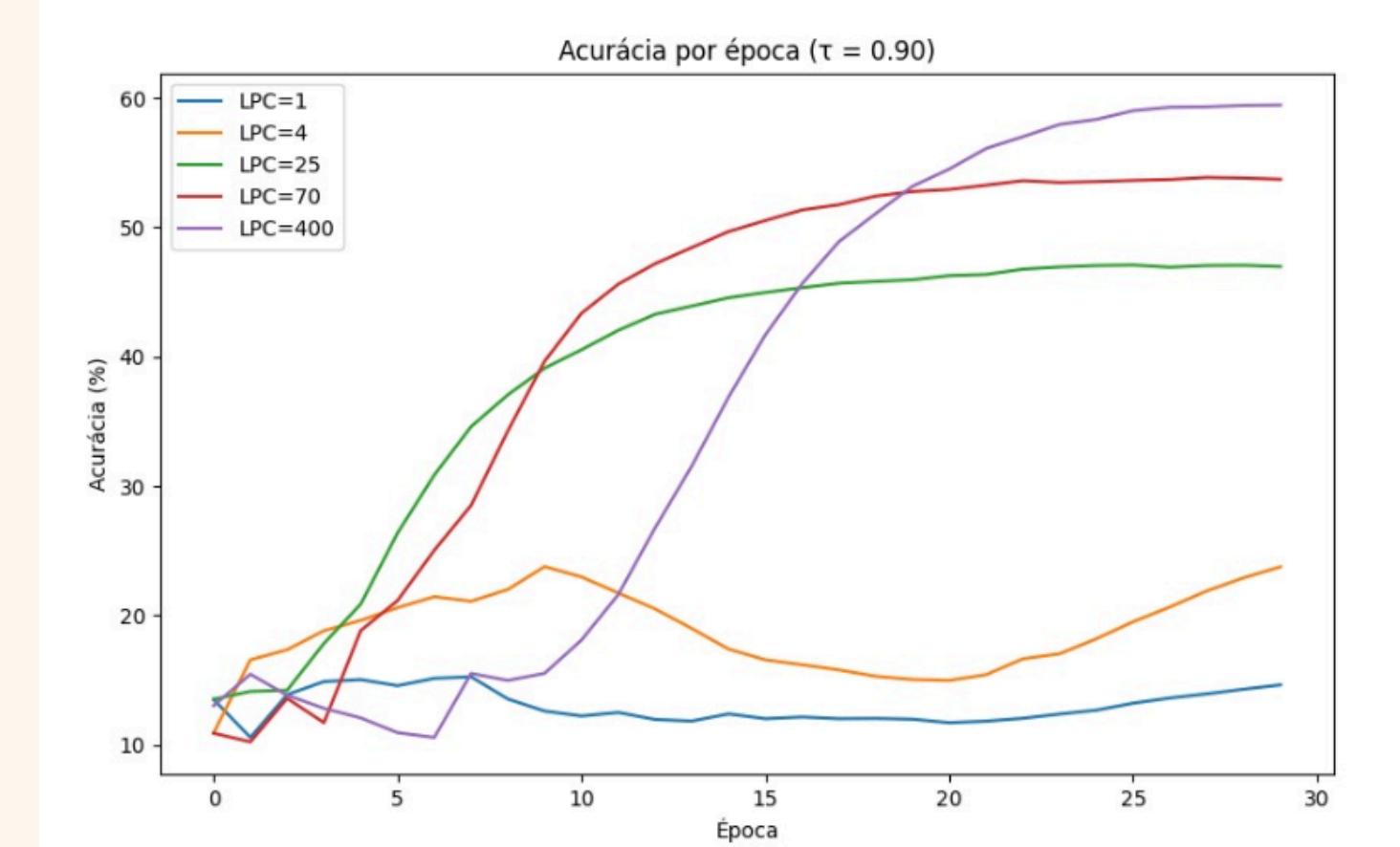
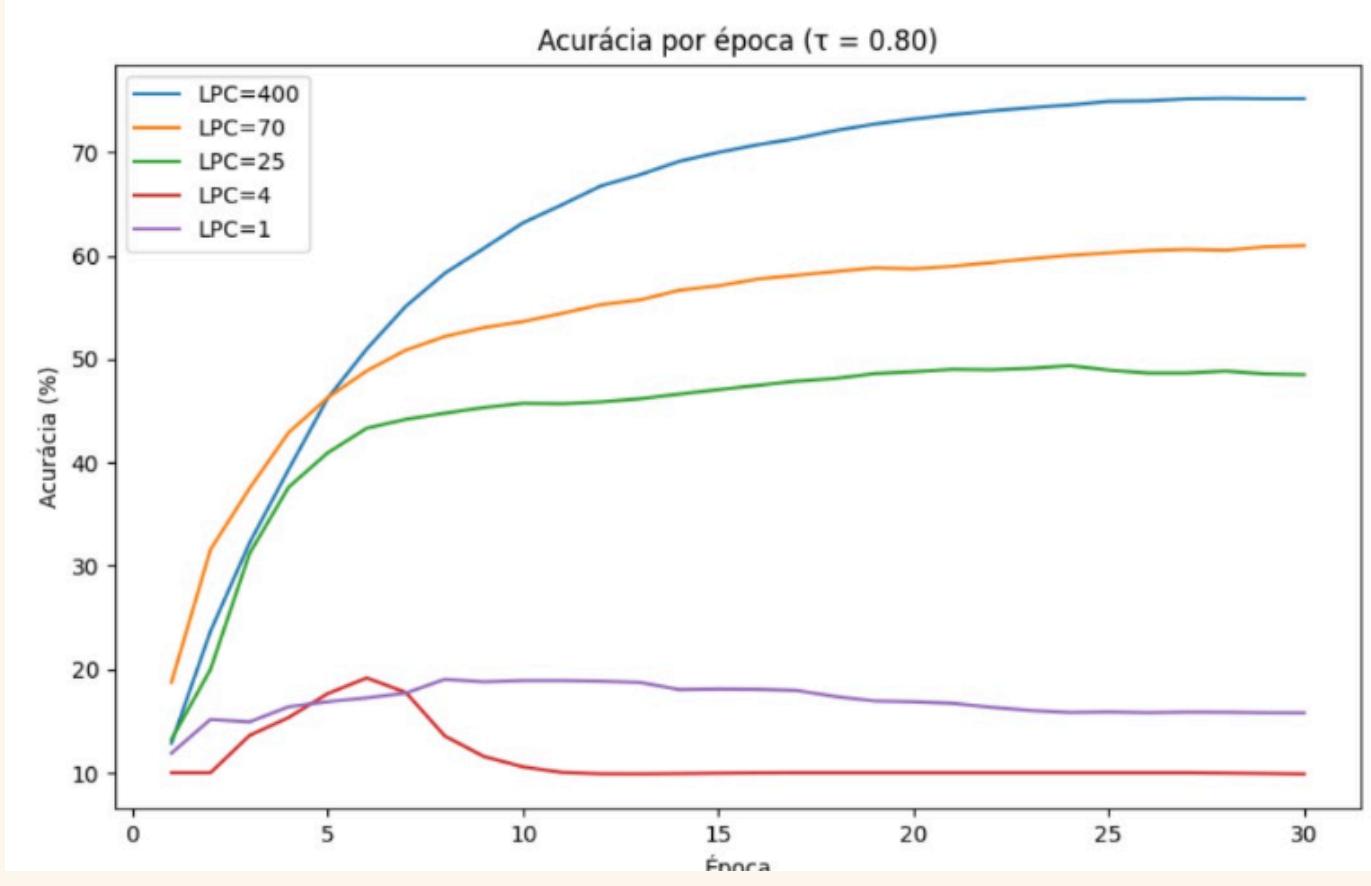
[Esperado] Ganhos maiores ao sair de $1 \rightarrow 4 \rightarrow 25$ rótulos/classe e ganhos marginais menores até 400, aproximando o supervisionado pleno.

[Esperado] τ mais alto aumenta a precisão dos pseudo-rótulos porém reduz cobertura; τ mais baixo amplia cobertura mas pode inserir ruído e derrubar a acurácia.

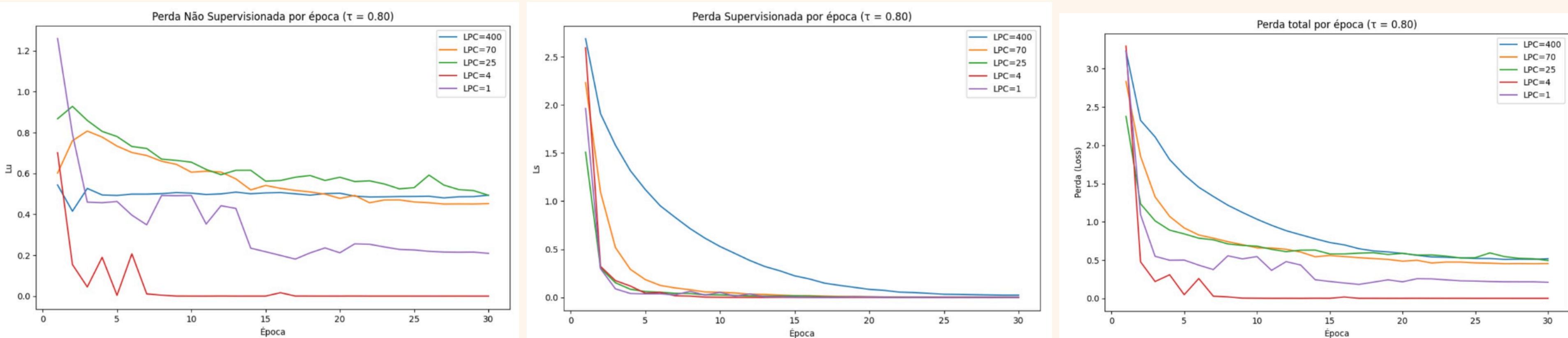
Qualidade dos pseudo-rótulos



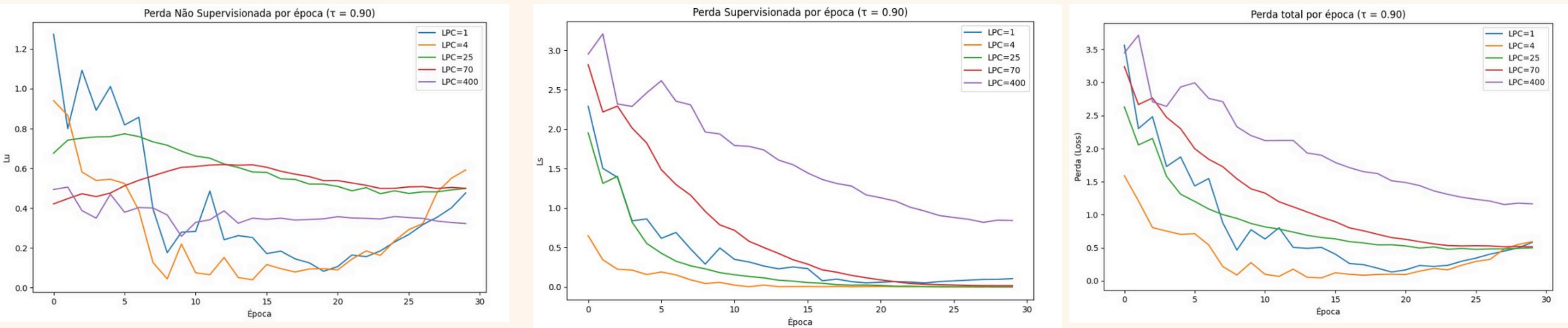
Impacto do threshold de confiança



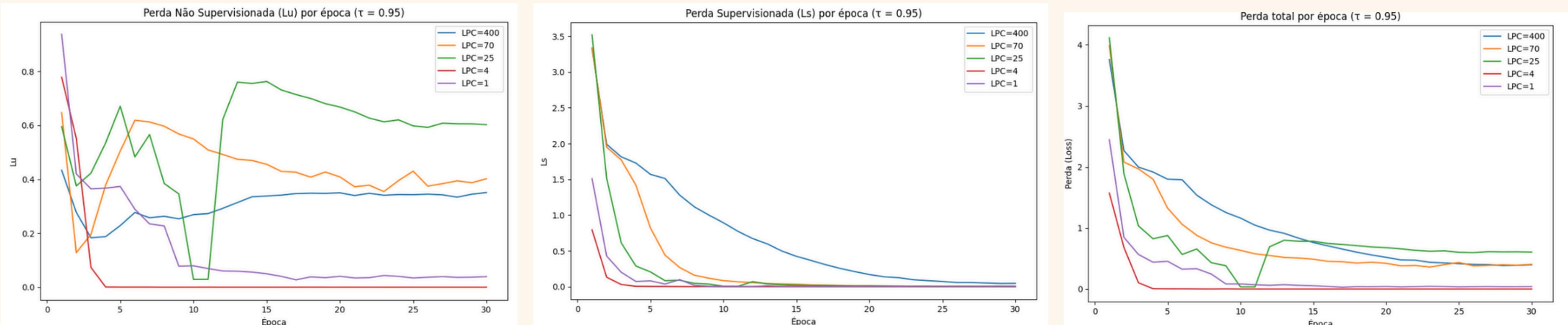
Impacto do augmentation e loss híbrida



Impacto do augmentation e loss híbrida



Impacto do augmentation e loss híbrida



Obrigado!