

Revisiting Neural Networks for Continual Learning: An Architectural Perspective

Aojun Lu¹, Tao Feng², Hangjie Yuan³, Xiaotian Song¹ and Yanan Sun^{1*}

¹Sichuan University

²Tsinghua University

³Zhejiang University

aojunlu@stu.scu.edu.cn, fengtao.hi@gmail.com, hj.yuan@zju.edu.cn
songxt@stu.scu.edu.cn, ysun@scu.edu.cn

<https://arxiv.org/pdf/2404.14829>

SIN5006 - INTELIGÊNCIA COMPUTACIONAL
ENTREGA PARCIAL

NILTON TADASHI ENTA - 12730911
VITOR CAMARGO PINHEIRO - 12693156

ArchCraft | Definição do problema

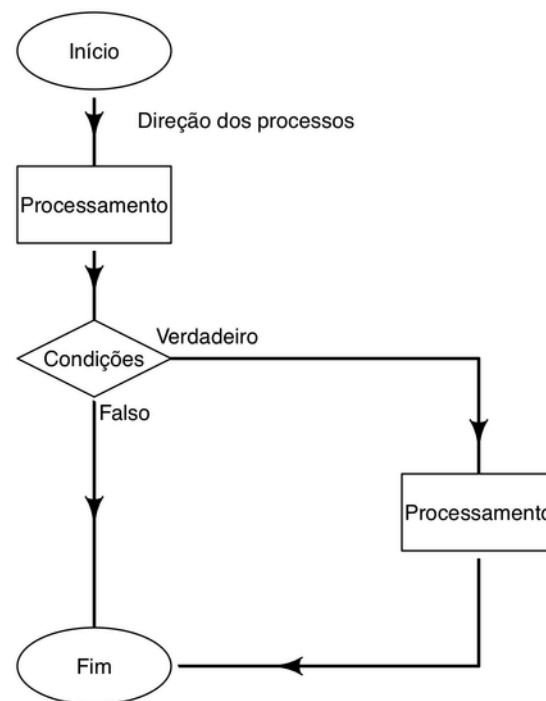
O Desafio:

- **Catastrophic Forgetting** na tarefa de Aprendizado Contínuo (CL)
- **CL refere-se ao incremento de novos dados no ciclo de vida de um modelo*

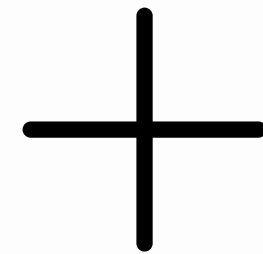
A Proposta:

- Um novo método de Neural Architecture Search (NAS), intitulado de **ArchCraft**. Focado em tarefas de CL para preencher a lacuna entre o **design da arquitetura de rede e o CL**.

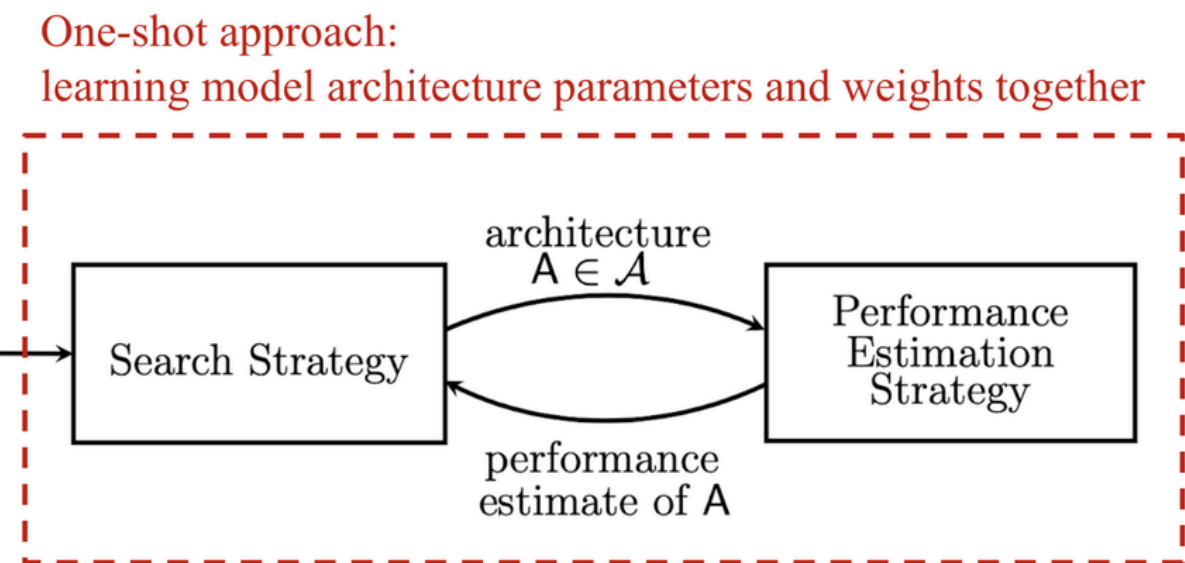
ArchCraft | Definição do Problema



Métodos Usuais

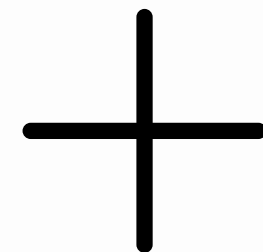


Search Space
 \mathcal{A}



ArchCraft

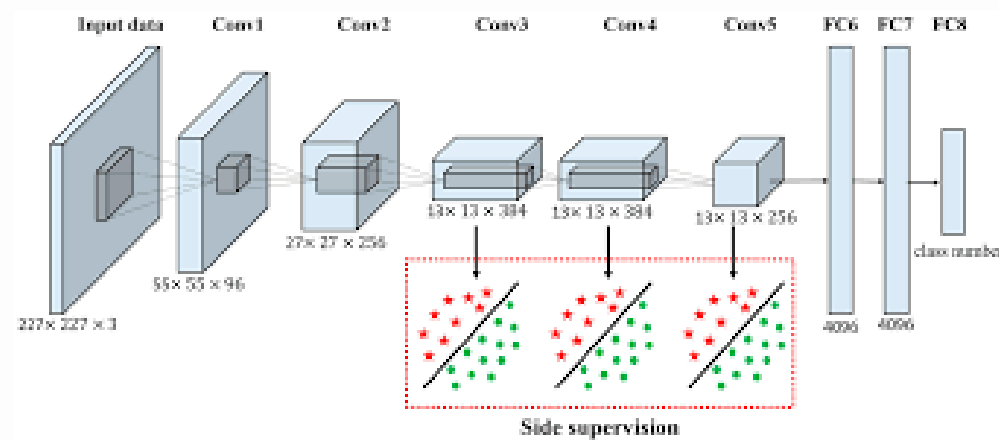
**Search Space
Especializado**



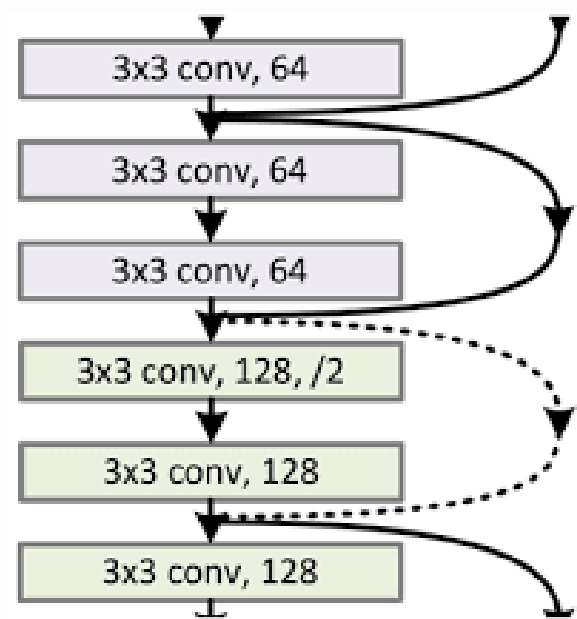
**Algoritmo
Genético**

ArchCraft | Metodologia

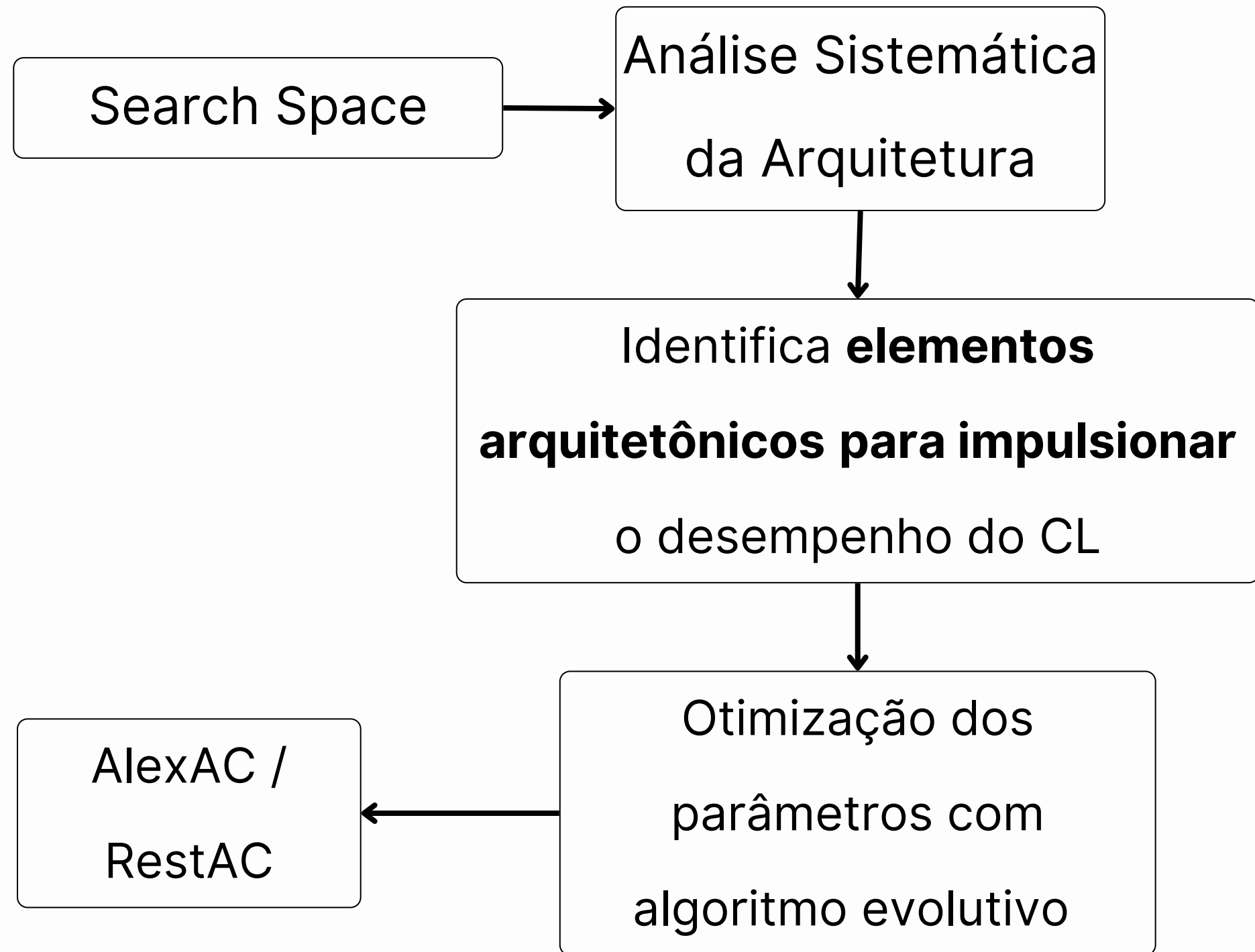
AlexNet



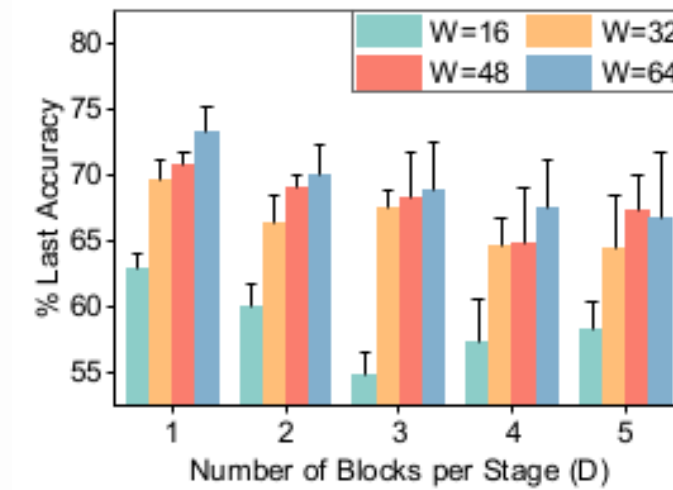
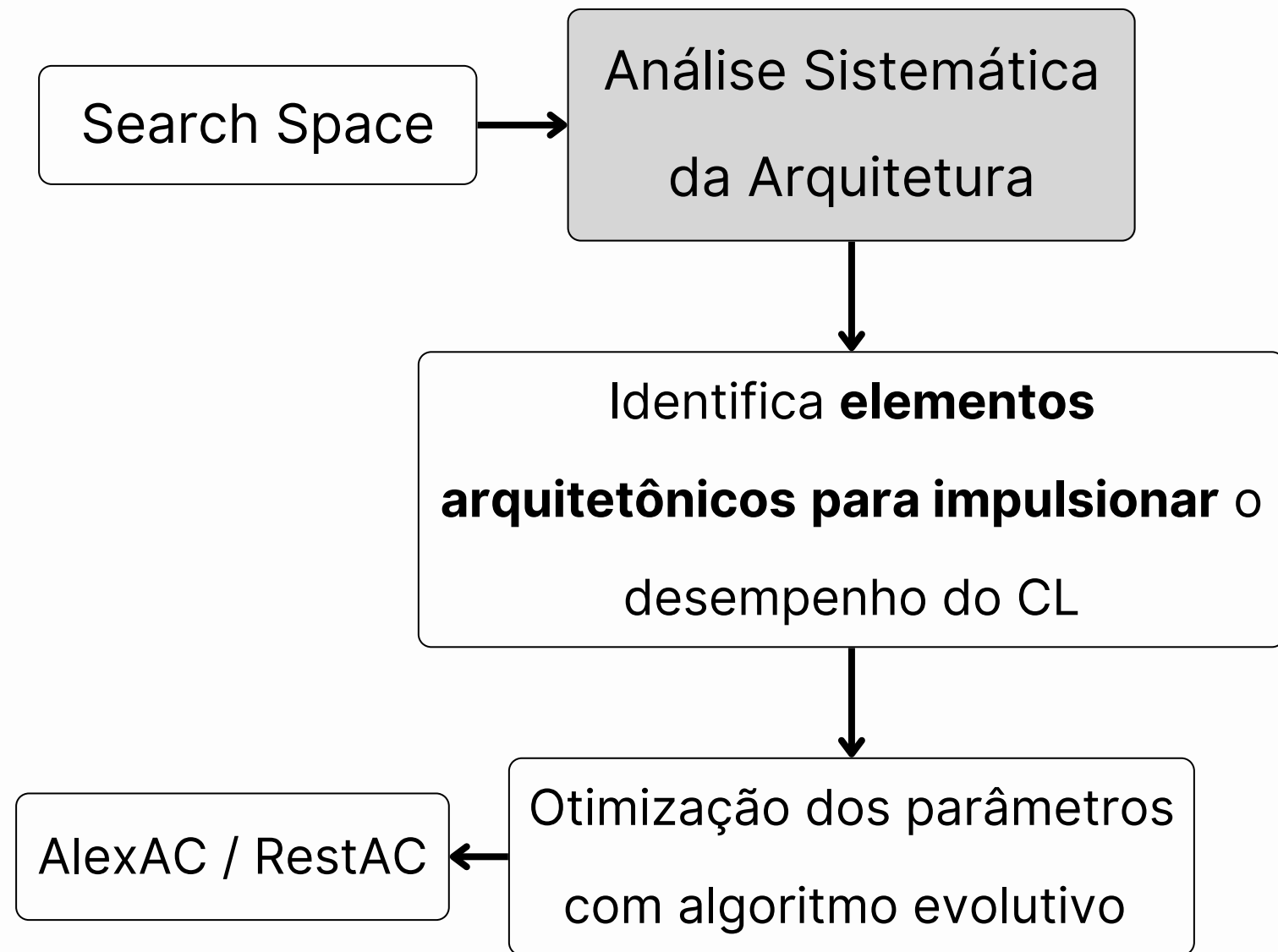
RestNet



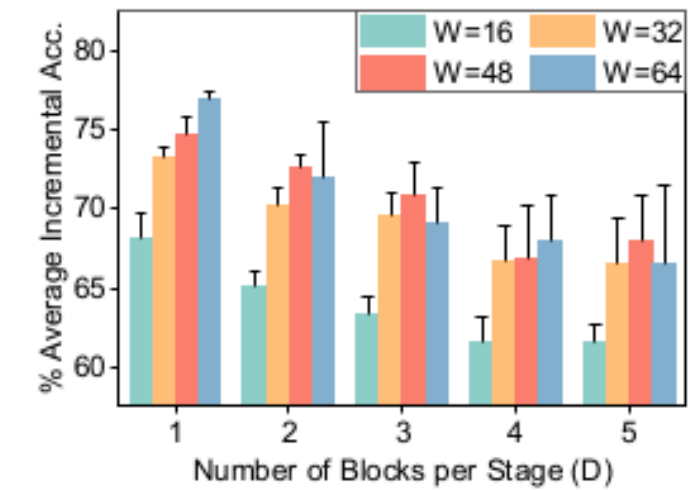
@51CTO博客



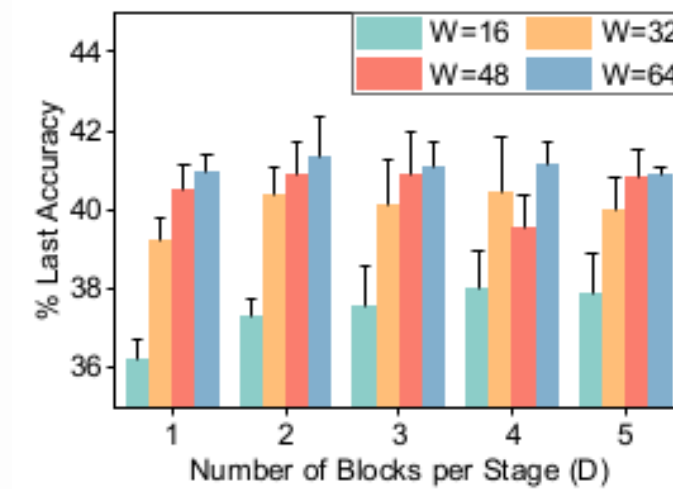
ArchCraft | Metodologia



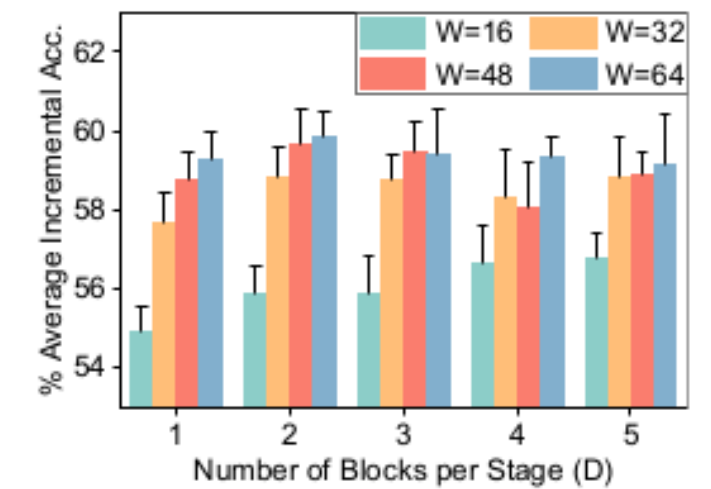
(a) LA in Task IL



(b) AIA in Task IL



(c) LA in Class IL



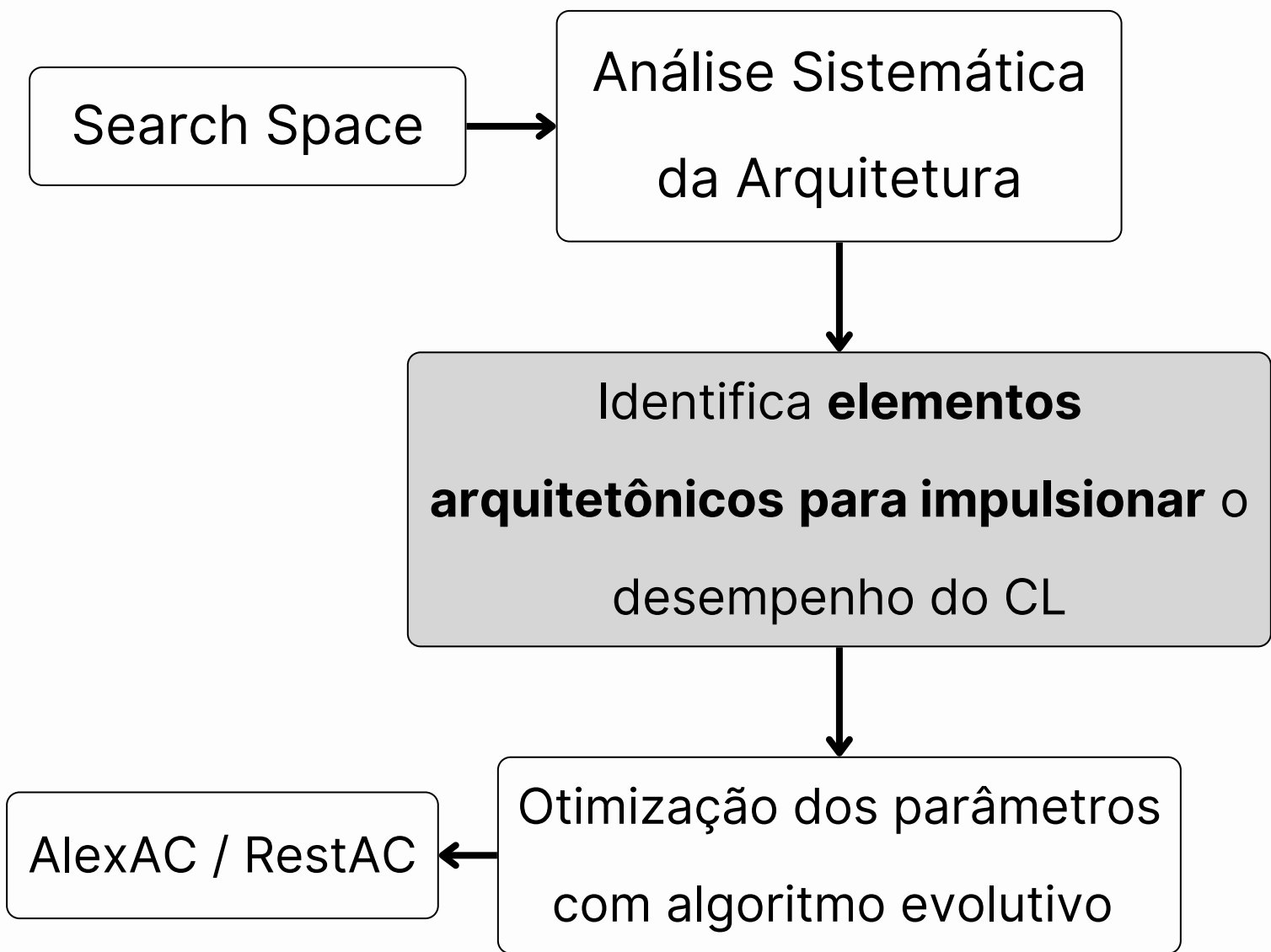
(d) AIA in Class IL

ArchCraft | Metodologia

Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24)

Network Components			Performance in Task IL				Performance in Class IL			
Down.	Skip	GAP	R32-LA	R32-AIA	R18-LA	R18-AIA	R32-LA	R32-AIA	R18-LA	R18-AIA
Strided Conv.	✓	✓	25.80±1.80	35.06±0.88	38.12±2.85	49.68±0.91	37.92±0.64	55.97±0.85	40.34±0.95	57.79±1.09
	✓	×	55.52±1.41	60.89±1.83	62.95±3.45	65.41±1.92	22.63±3.36	27.37±11.18	38.85±1.60	50.29±3.01
	×	✓	25.29±1.48	31.74±0.90	30.33±2.28	41.75±0.89	30.74±2.11	46.99±1.48	38.45±0.66	56.86±0.82
	×	×	38.30±1.17	45.58±1.76	57.66±1.72	62.13±1.02	27.63±2.79	38.66±6.24	33.99±3.77	45.98±4.53
Max Pooling	✓	✓	25.91±2.35	35.58±0.78	39.89±1.80	53.15±0.69	38.27 ±0.88	56.79 ±0.88	40.50 ±0.34	59.53 ±1.26
	✓	×	57.31 ±1.56	63.35 ±1.65	63.94 ±3.13	68.74 ±3.27	24.91±3.58	22.23±13.19	40.00±0.67	54.11±1.79
	×	✓	24.54±0.53	32.16±0.87	30.46±2.52	42.24±1.08	30.69±1.27	47.50±1.42	37.34±0.72	56.53±0.83
	×	×	36.53±0.85	44.72±1.31	60.16±0.88	65.41±0.69	16.67±13.00	25.64±18.06	33.92±3.98	47.83±4.96
Avg Pooling	✓	✓	24.37±1.72	35.00±0.92	38.47±2.01	53.52±0.67	38.25±0.52	56.67±0.61	39.85±1.52	57.53±1.14
	✓	×	56.16±2.93	62.60±1.74	63.92±3.22	68.56±2.78	27.62±5.27	34.69±11.64	37.35±3.96	49.74±4.24
	×	✓	23.70±1.92	31.50±0.88	30.21±1.27	42.82±0.80	30.33±2.09	46.55±1.24	37.02±0.92	55.99±0.48
	×	×	35.86±2.09	44.93±2.42	60.68±0.86	65.78±1.29	23.04±11.10	33.05±14.94	34.84±1.67	47.63±1.41

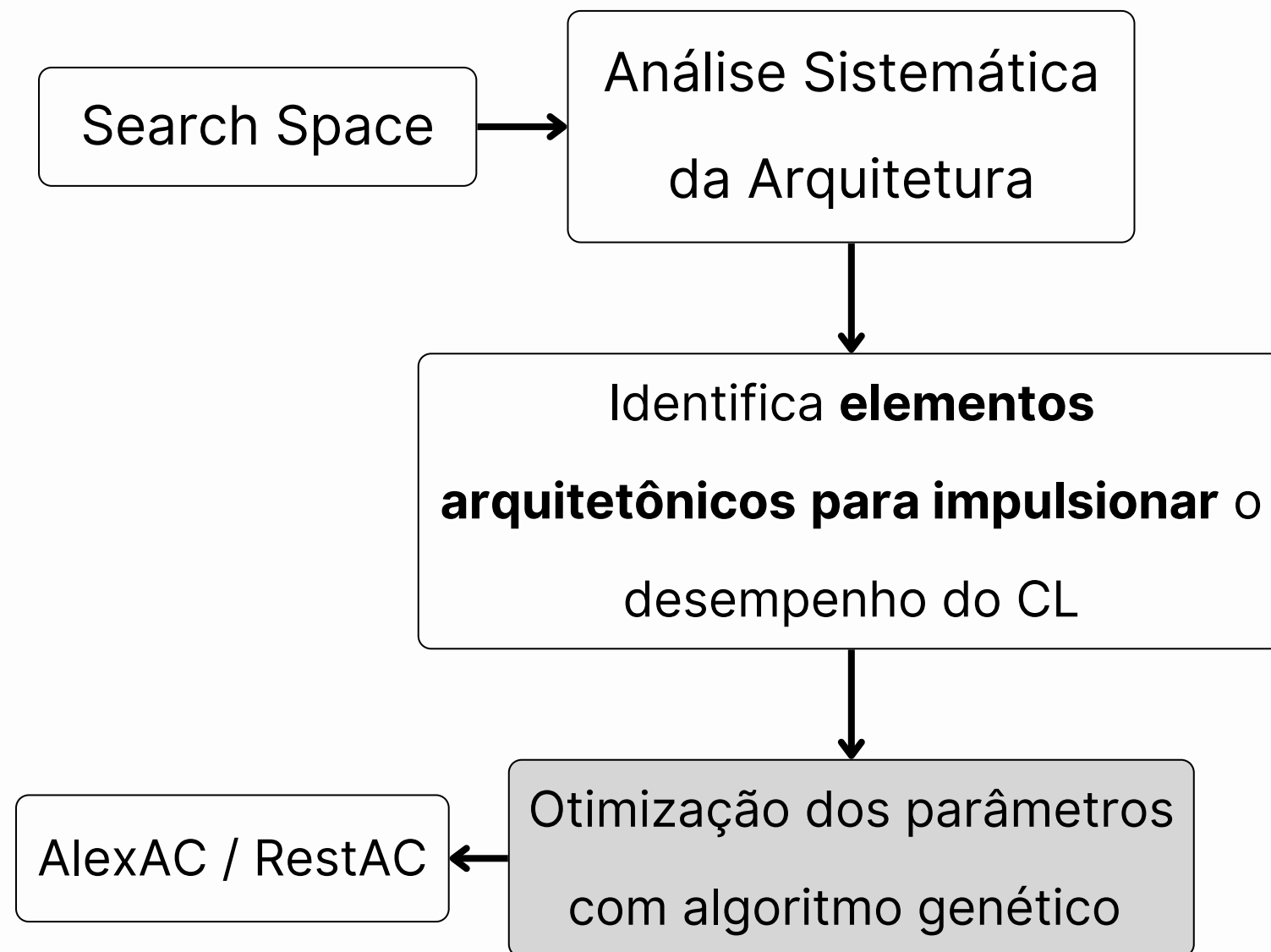
ArchCraft | Metodologia



Initial Width	Final Width	Performance in Task IL		Performance in Class IL	
		LA	AIA	LA	AIA
16	64 256	59.48 \pm 2.03	65.65 \pm 0.89	35.82 \pm 0.55	54.43 \pm 0.47
		71.51 \pm 0.87	73.89 \pm 0.80	36.16 \pm 0.52	55.18 \pm 0.63
32	128 256	68.24 \pm 0.72	71.74 \pm 0.69	39.44 \pm 0.81	57.83 \pm 0.72
		74.18 \pm 0.94	76.18 \pm 0.46	40.02 \pm 0.74	58.02 \pm 0.87
48	192 256	70.96 \pm 1.70	74.94 \pm 0.80	40.52 \pm 0.92	58.93 \pm 0.77
		73.12 \pm 1.46	76.64 \pm 0.63	40.53 \pm 0.75	59.08 \pm 0.67
64	256	73.14 \pm 1.51	76.61 \pm 0.66	40.97 \pm 0.56	59.48 \pm 0.85

Table 2: The CL performance on different width configurations.

ArchCraft | Metodologia



Novos **indivíduos (descendentes)** são criados em um algoritmo evolutivo.

Seleção dos pais:

- Dois indivíduos (soluções) são escolhidos aleatoriamente da população atual.
- O que tiver melhor desempenho (fitness) é escolhido para ser o "pai".

Mutação:

- O pai é copiado para criar um descendente.

Em seguida, uma parte do código desse descendente é alterada aleatoriamente (mutação).

ArchCraft | Metodologia

Algoritmo Genético:

old_code \leftarrow [largura, profundidade, pooling[2, 4, 6, 6, 6], double[0, 0, 6, 6, 6]]

WHILE parâmetros < 1 milhão **AND** tentativas < 101:

old_code \leftarrow mutação $\frac{1}{6}$ de probabilidade

largura \leftarrow largura * (ou "/") random == mod(4) e mínimo 8

profundidade \leftarrow profundidade * (ou "/") random

pooling, double \leftarrow ajuste proporcional

Executa 20 iterações de 5 mutações consecutivas

ArchCraft | Medidas de Acurácia para o CL

Parâmetros de Treino:

SGD: Stochastic Gradient Descent

SGD (learning rate: 0.01,
momentum: 0.9,
weight decay: 0.0002,
maximize: False,
differentiable: False)

Parâmetros de Treino:

Acurácia Incremental Média (AIA):

- Média das acurácias por classe

Catastrophic Forgetting:

- Média da diferença entre o máximo valor

PSEUDOCODIGO

DEFINE VARIÁVEIS LOCAIS

incremento <- 50

individuo_id <- 0

code <- [profundidade, largura, pool[1, ..., 5], double[1, ..., 5]]

chosen_network <- 'arch_craft'

Algoritmo TrainModel:

grad_clip <- 10

epoch <- 2

learning_rate <- 0.01

Carrega os dados CIFAR-100 divididos em tarefas incrementais

Se rede == 'arch_craft': importa e instancia arquitetura Net personalizada

Se rede == 'alexnet': importa e instancia versão AlexNet

Mostra o número total de parâmetros do modelo

Inicializa o treinamento com otimizador SGD (Gradiente Descendente Estocástico)

Imprime a função de perda e o otimizador

Inicializa matrizes para armazenar acurácias, perdas, e medidas por tarefa

for tarefa **in** classes:

Treina o modelo

Salva os resultados

for conjunto **in** range(tarefas já processadas):

Testa a conjunto

Salva os resultados

Calcula a acurácia média após a tarefa

Calcula forgetting <- $\text{sum}(\text{acurácia máxima} - t[x]) / \text{len}(t)$

Imprime matriz de acurácia

Calcula e imprime métrica média ap (AIA)

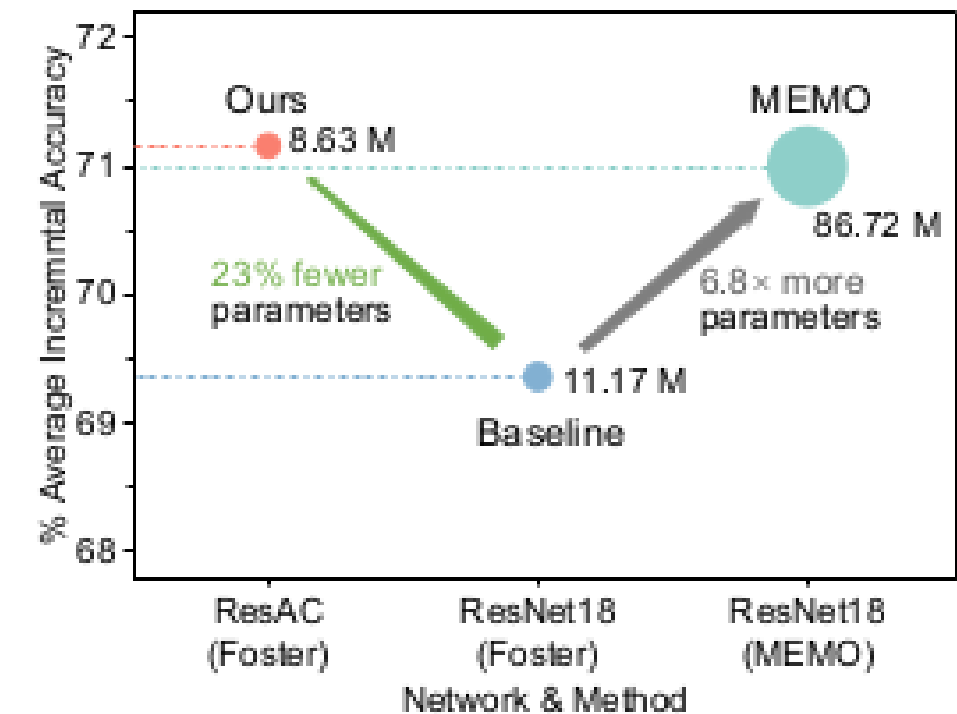
Calcula e imprime acurácia

Retorna a acurácia final

ArchCraft | Resultados

Por que a quantidade de parâmetros é importante?

- Capacidade de Representação (Expressividade)
 - Mais parâmetros → maior capacidade de modelar funções complexas.
 - Permite que a rede aprenda padrões detalhados e relações não lineares.
- Risco de Overfitting
- Capacidade de Generalização
 - Modelos com parâmetros suficientes, mas não excessivos.
- Custo Computacional e Memória
 - Mais parâmetros → mais memória e tempo de computação, tanto para treino quanto para inferência.
- Escalabilidade e Eficiência
 - Modelos com muitos parâmetros podem ser difíceis de treinar ou ajustar com dados contínuos (aprendizado online)



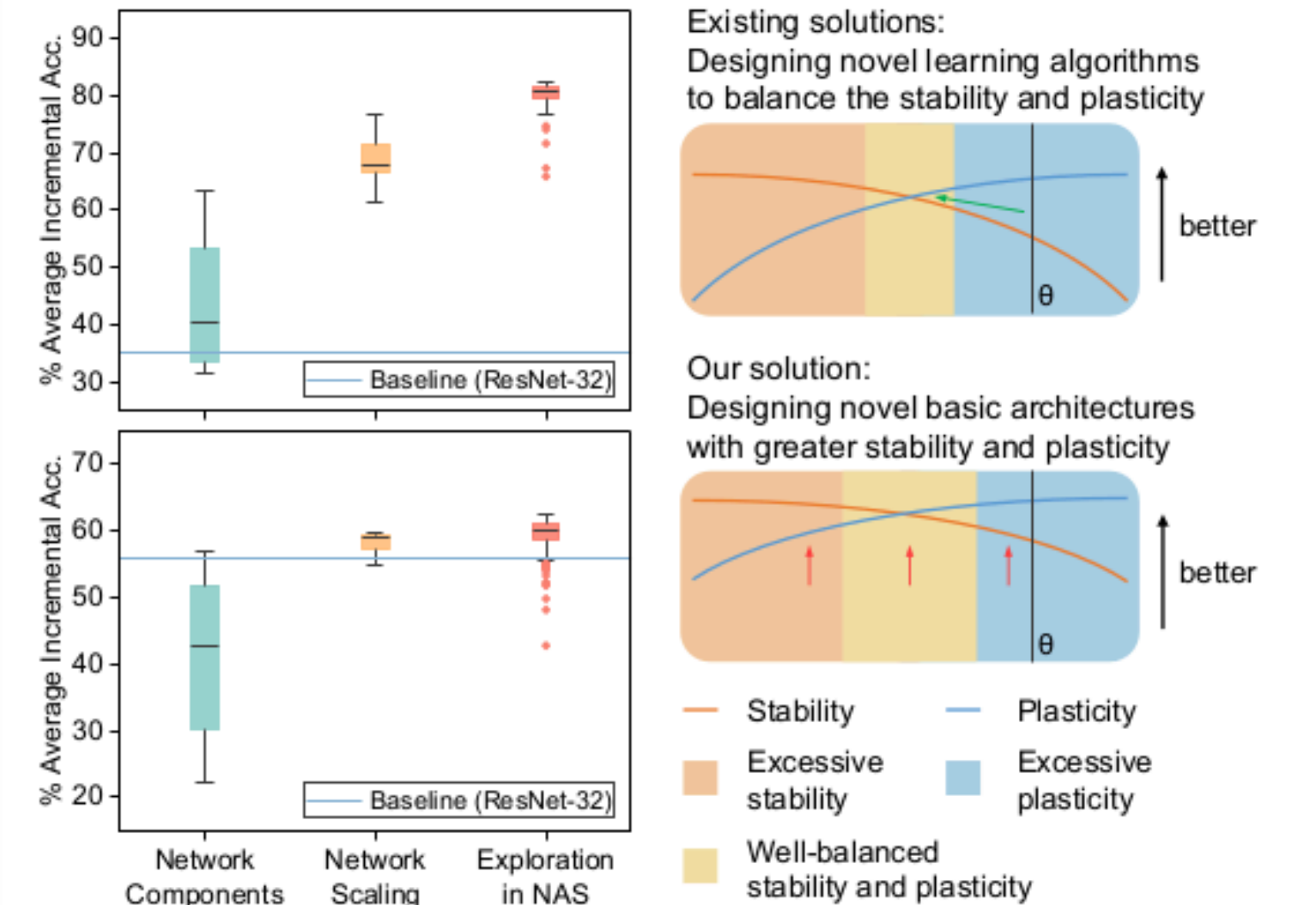
ArchCraft | Resultados

Equilíbrio entre Plasticidade e Estabilidade

Um dos principais desafios em aprendizado contínuo é equilibrar plasticidade com estabilidade:

- Plasticidade demais: o modelo aprende novas tarefas, mas esquece as antigas (catastrophic forgetting).

Estabilidade demais: o modelo preserva o que já sabe, mas não aprende bem novas informações (underfitting nas novas tarefas).



Network	#P (M)	Method	I100-inc5	I100-inc10
AlexNet	6.71	SGD	35.36	33.38
AlexAC-A	6.28↓ 6%	SGD	52.02	45.86

Table 4: The last accuracy on Imagenet-100 (I100) in *Task IL*.

ArchCraft | Referências

LU, Aojun; FENG, Tao; YUAN, Hangjie; SONG, Xiaotian; SUN, Yanan. Revisiting neural networks for continual learning: an architectural perspective. arXiv:2404.14829, 2024. Disponível em: <https://arxiv.org/abs/2404.14829>. Acesso em: 30 abr. 2025.

ArchCraft | Referências

LU, Aojun; FENG, Tao; YUAN, Hangjie; SONG, Xiaotian; SUN, Yanan. Revisiting neural networks for continual learning: an architectural perspective. arXiv:2404.14829, 2024. Disponível em: <https://arxiv.org/abs/2404.14829>. Acesso em: 30 abr. 2025.

ArchCraft | Metodologia

Base de Dados:

- CIFAR100 (*100 classes*)

Representação do CL:

- Incremento em lotes de mesmo tamanho
- As novas classes são treinadas com $\frac{1}{3}$ das épocas do treino inicial

Search Space:

- AlexNet
- ResNet

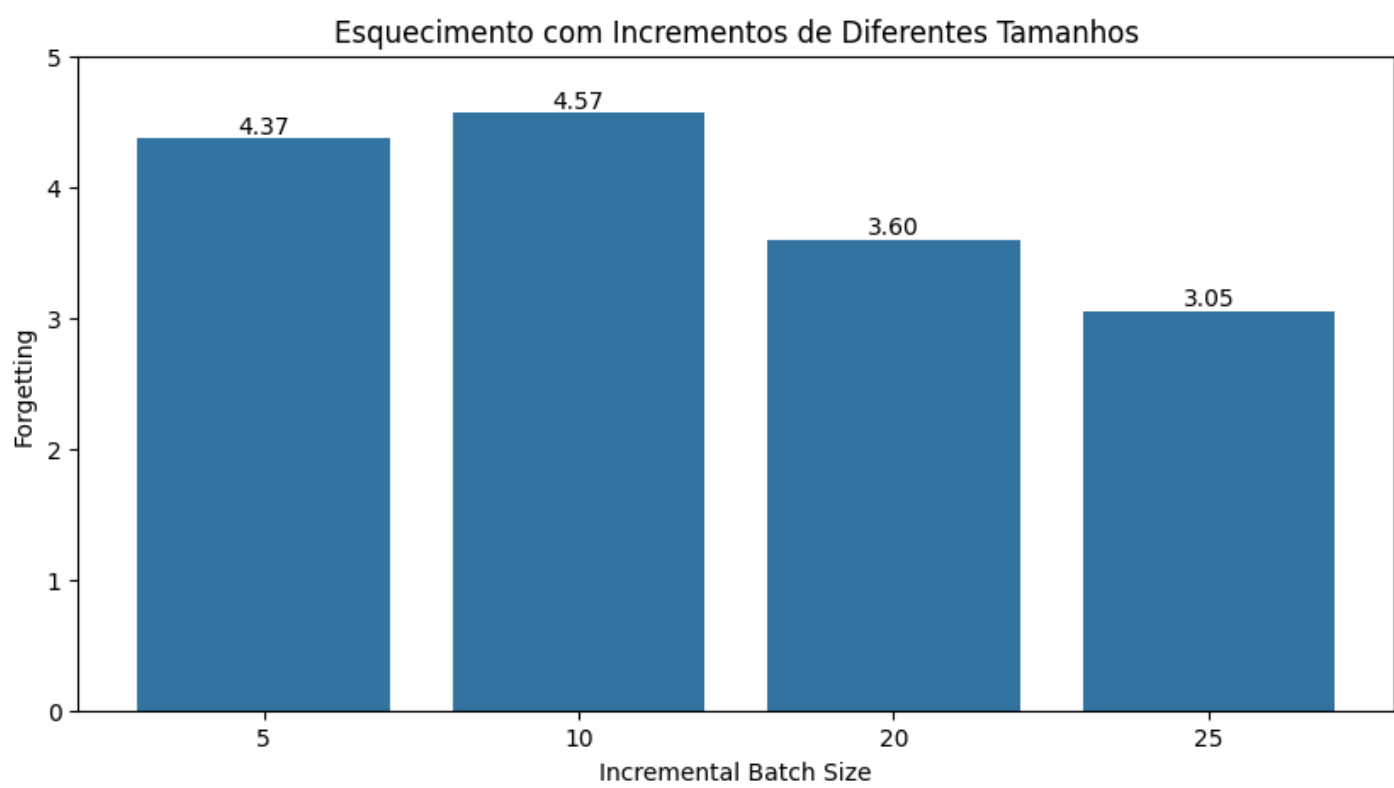
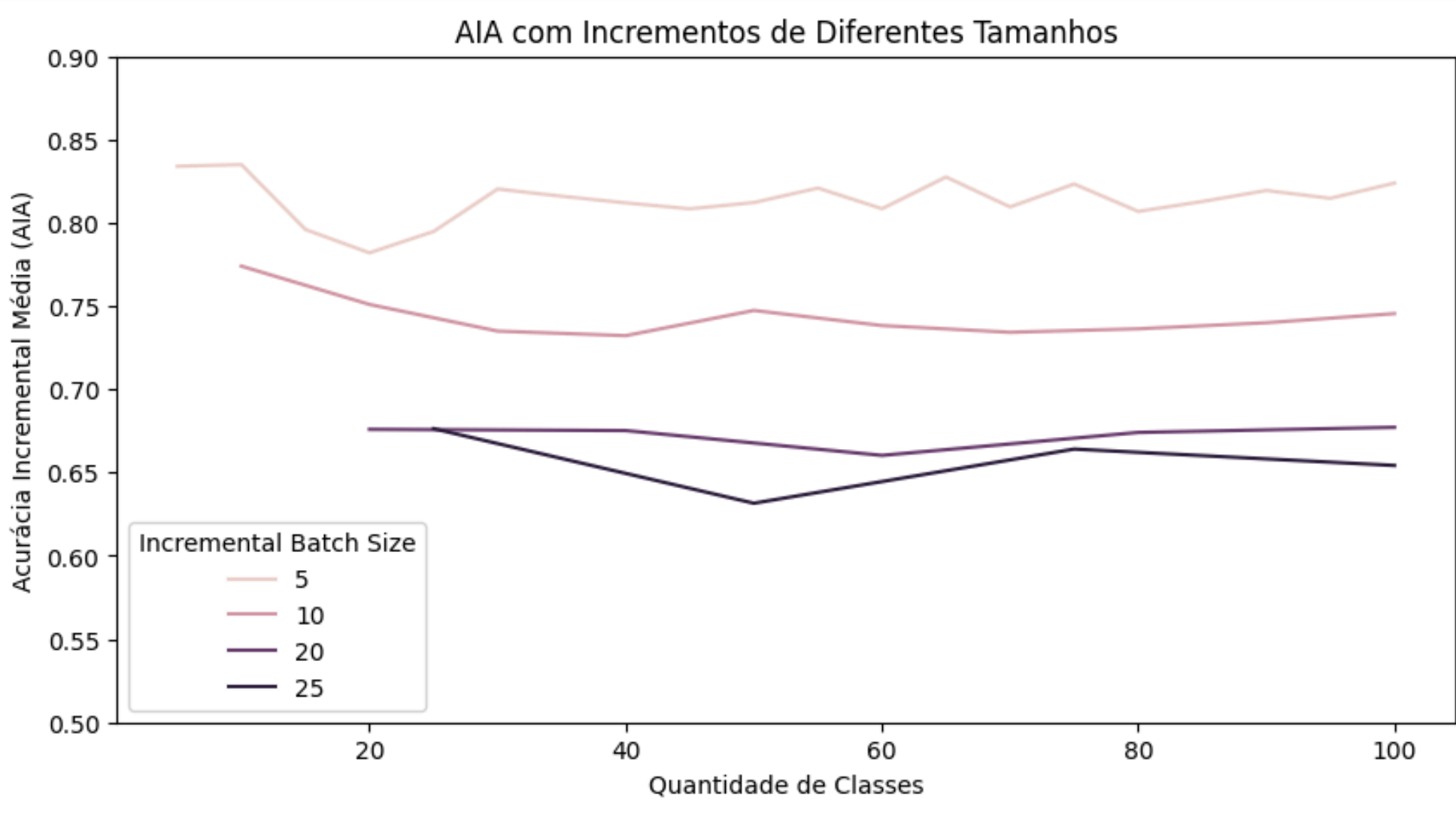
Algoritmo Genético:

- Mutações estocásticas nos parâmetros selecionados da arquitetura
- Mantém a complexidade em até 1 milhão de parâmetros

Medidas de Avaliação:

- Acurácia Incremental Média (AIA)
- Catastrophic Forgetting

ArchCraft | Resultados Iniciais



ANEXOS

ArchCraft | Métodos (Visão Geral)

ArchCraft é um novo método de Neural Architecture Search (NAS).

Focado em Aprendizado Contínuo (CL) para preencher a lacuna entre o **design da arquitetura de rede e o CL**.

I. Search Space: Espaço de busca personalizado obtido a partir de **busca empírica** [1].

II. Friendly Search Strategy: Uma estratégia baseada em algoritmos genéticos para explorar o Search Space [2].

[1] O ArchCraft ele é projetado com base em **observações empíricas sobre seus componentes arquitetônicos**.

O experimento consistiu com uma ResNet. Partindo de configurações consideradas ótimas para essas tarefas (Incremental Learning (IL)): Max Pooling, Skip connection e Global Avg. Pooling

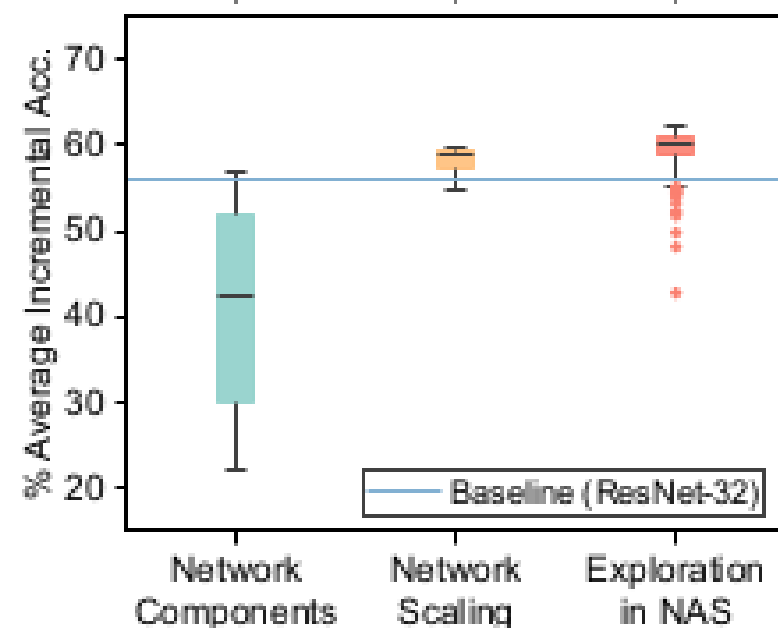
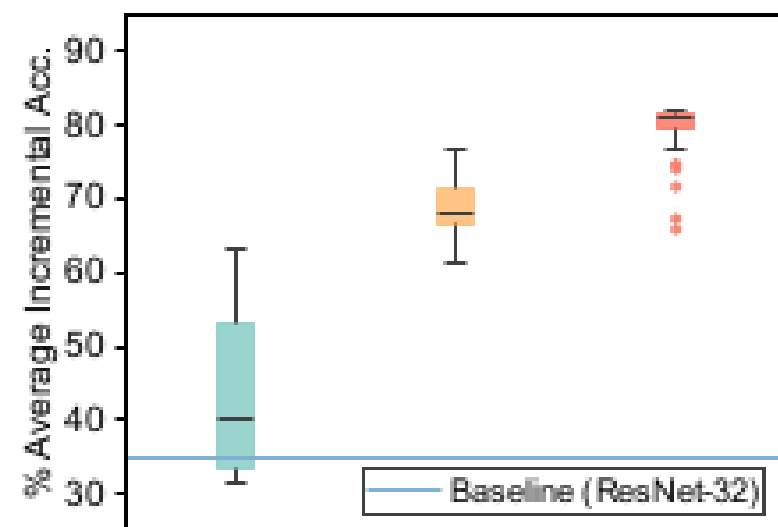
[2] **Inicialização:** Uma população de arquiteturas é gerada aleatoriamente

Avaliação da performance: durante as iterações evolutivas os indivíduos são avaliados por Acurácia Incremental Média (AIA).

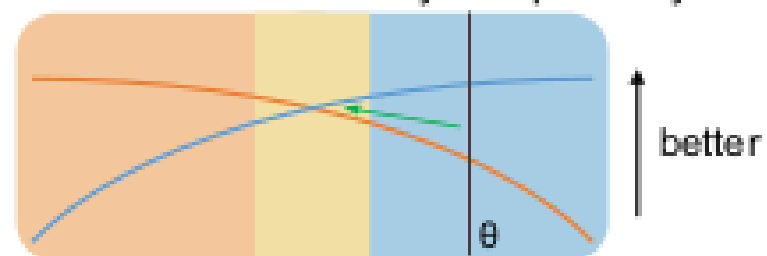
Geração de novos indivíduos: novas arquiteturas são geradas a partir de dois indivíduos, escolhidos aleatoriamente, chamados de filho. O filho é resultado da mutação dessas duas arquiteturas.

Nova população: É gerada a partir dos filhos com melhor aptidão.

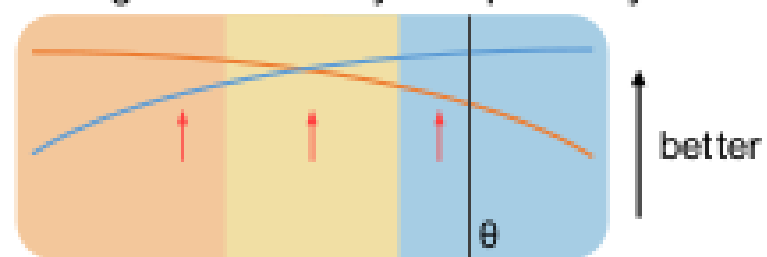
Anexos | Resultados



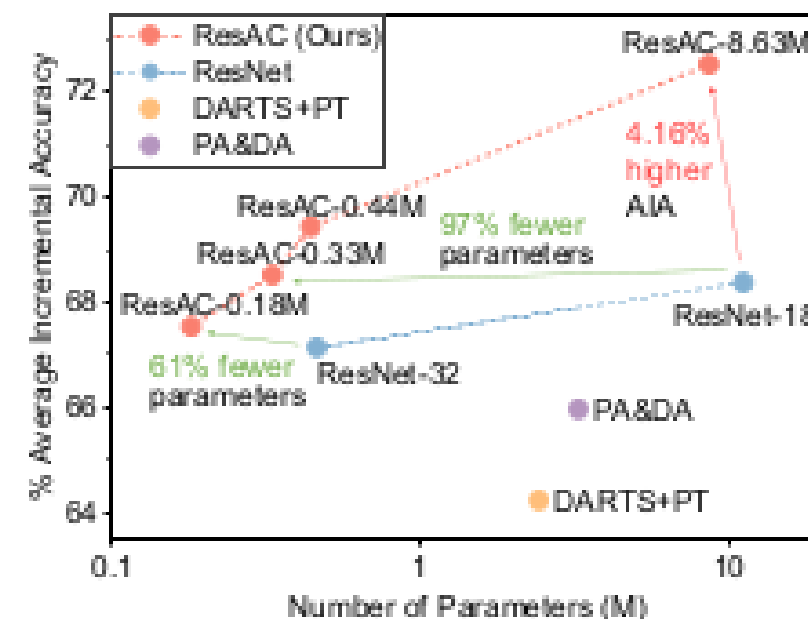
Existing solutions:
Designing novel learning algorithms
to balance the stability and plasticity



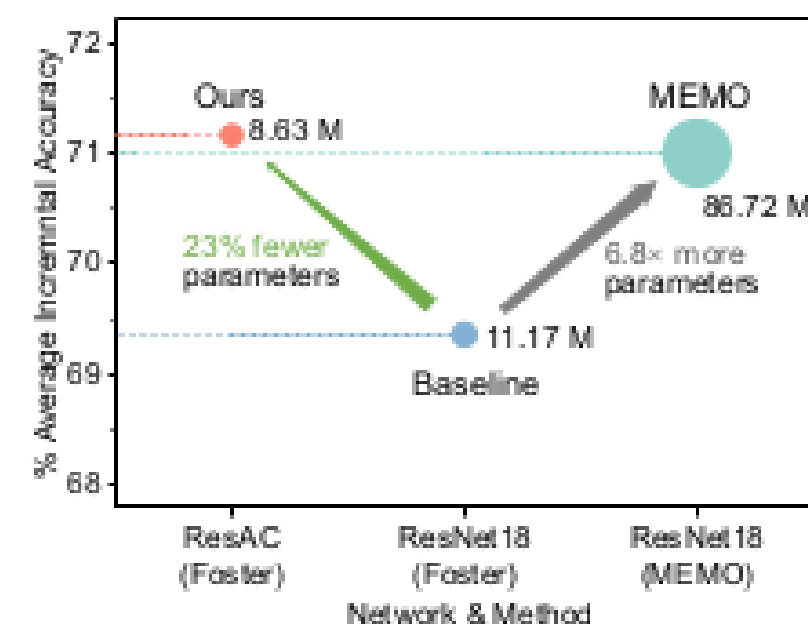
Our solution:
Designing novel basic architectures
with greater stability and plasticity



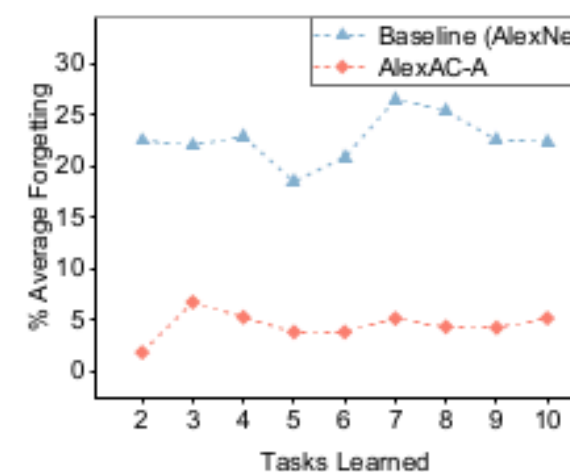
- Stability
- Excessive stability
- Well-balanced stability and plasticity
- Plasticity
- Excessive plasticity



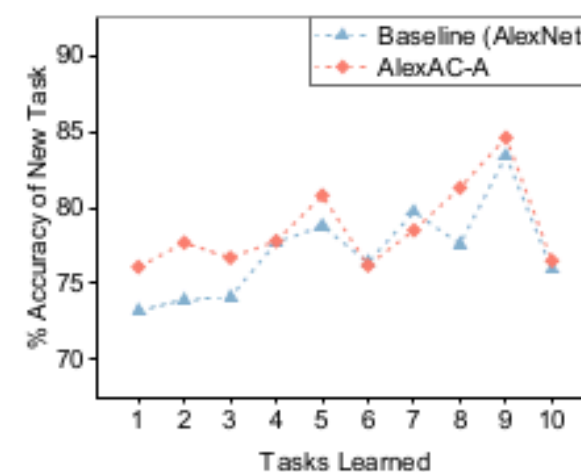
(a)



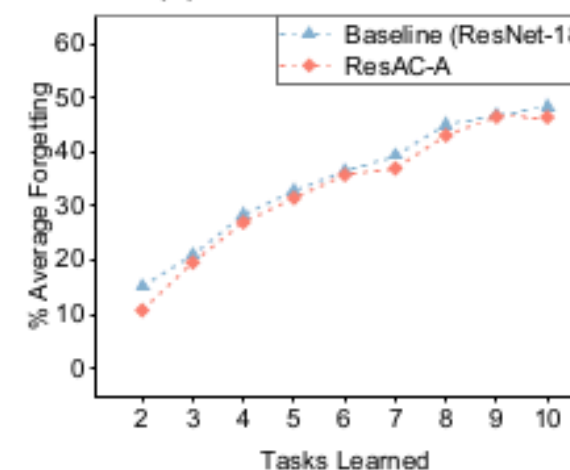
(b)



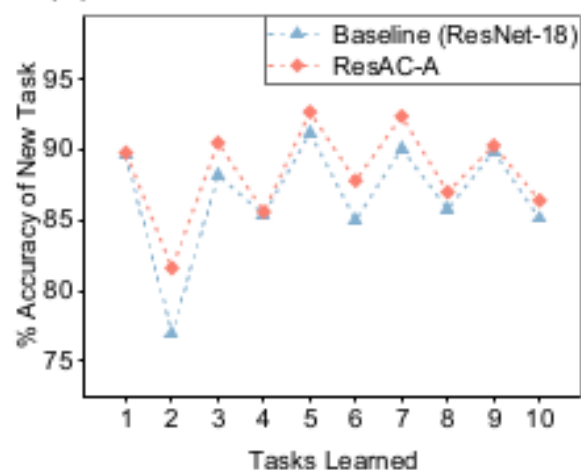
(a) AF in Task IL



(b) New Task Acc. in Task IL



(c) AF in Class IL



(d) New Task Acc. in Class IL

Anexos | Resultados

Method	Network	#P (M)	C100-inc5		C100-inc10		I100-inc5		I100-inc10		Max Improvement	
			LA	AIA	LA	AIA	LA	AIA	LA	AIA	LA	AIA
Replay	ResNet-32	0.46	39.10	58.17	40.02	58.21	-	-	-	-	-	-
	ResAC-B	0.44(↓ 4%)	40.45	59.67	42.79	59.99	-	-	-	-	+2.77	+1.78
	ResNet-18	11.17	40.04	58.80	43.23	60.42	36.30	57.30	41.00	59.21	-	-
	ResAC-A	8.63↓ 23%	42.99	62.52	46.62	63.36	36.78	57.40	42.44	60.07	+3.39	+3.72
iCaRL	ResNet-32	0.46	46.67	63.47	48.80	64.18	-	-	-	-	-	-
	ResAC-B	0.44↓ 4%	47.94	64.17	50.11	64.42	-	-	-	-	+1.31	+0.70
	ResNet-18	11.17	47.32	64.13	52.77	66.04	44.10	62.36	50.98	67.11	-	-
	ResAC-A	8.63↓ 23%	52.6	68.71	55.52	69.62	45.12	63.98	52.46	68.42	+5.28	+4.58
WA	ResNet-32	0.46	46.95	62.93	53.35	66.61	-	-	-	-	-	-
	ResAC-B	0.44↓ 4%	51.31	66.39	54.89	67.73	-	-	-	-	+4.36	+3.46
	ResNet-18	11.17	45.11	62.06	56.59	68.89	46.06	62.96	55.04	68.60	-	-
	ResAC-A	8.63↓ 23%	53.23	69.19	59.79	71.40	49.94	67.20	58.86	71.56	+8.12	+7.13
Foster	ResNet-32	0.46	47.78	62.36	54.36	67.14	-	-	-	-	-	-
	ResAC-B	0.44↓ 4%	53.50	67.34	58.17	69.44	-	-	-	-	+5.72	+4.98
	ResNet-18	11.17	49.03	61.97	55.98	68.38	53.26	65.20	60.58	69.36	-	-
	ResAC-A	8.63↓ 23%	57.22	69.99	61.44	72.54	54.32	66.41	61.94	71.16	+8.19	+8.02

Table 5: The CL performance of ArchCraft in *Class IL*. ‘#P’ represents the number of parameters of the network used.