Make Continual Learning Stronger via C-Flat

🍜 ♪: C$_b$-a pitch that is one semitone lower than C tune.

C-Flat (♪)
Motivation

Characterizing generalization from loss landscapes to promote CL leads to CL that are:

- **A less-than-stellar performance**.

- **Not exhaustively explored on loss landscape.**
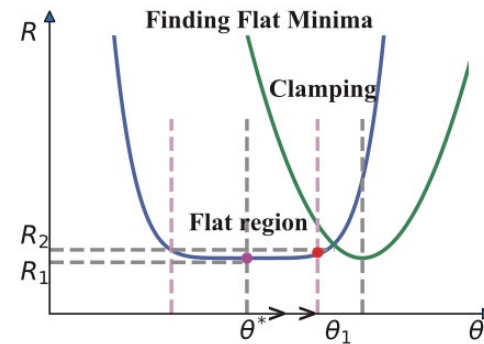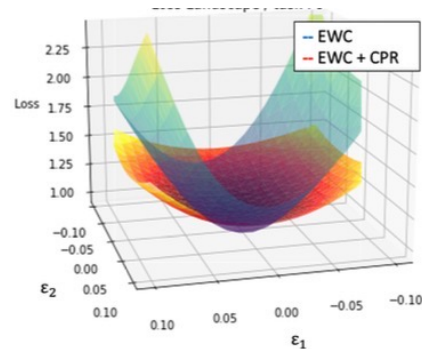
Seeking for flat minima (Sharpness-aware Minimization) has proven to be a strong training regime for continual learning.

Our hypothesis

- **Featuring a flatter loss landscape** upon sequential arriving tasks can overcome forgetting, thereby ensure CL stability.

[NeurIPS 2024] "Make Continual Learning Stronger via C-Flat". 🥚 ♪: **C_b-a pitch that is one semitone lower than C tune.**

C-Flat (♪)
Challenge

Our hypothesis

- **Featuring a flatter loss landscape** upon sequential arriving tasks can overcome forgetting, thereby ensure CL stability.

Contribution

- We **propose C**ontinual **Flat**ness **(C-Flat)** optimization to **Make Continual Learning Stronger** .
- We **propose a unified framwork** of C-Flat covering divers CL method categories, and prove that **Flatter is Better** in early all cases.
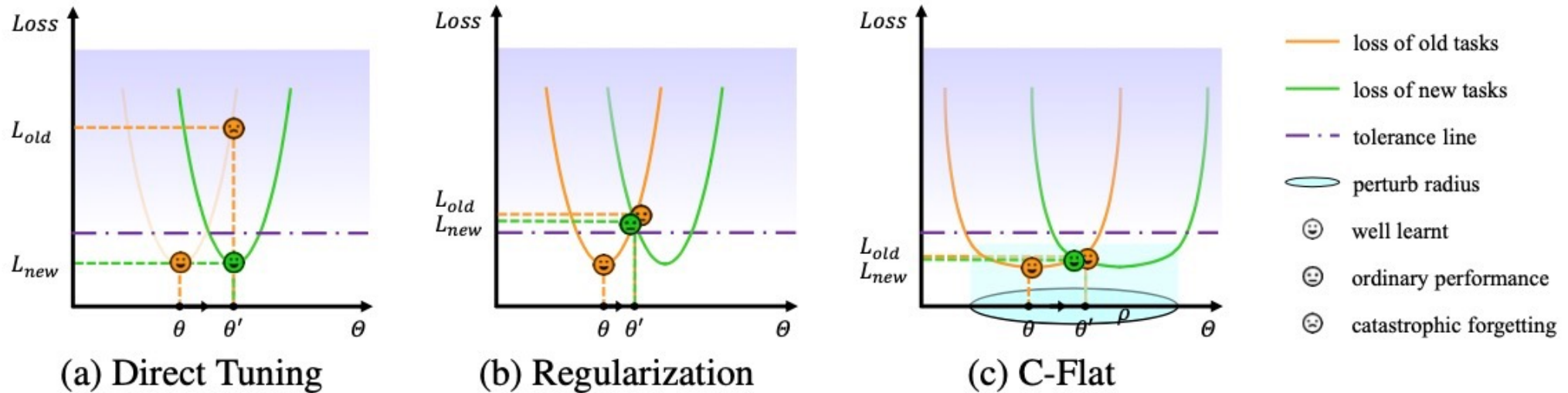
```
1  from  ...  import  C_Flat
2
3  # initialize optimizer for CL.
4  C_Flat_optimizer = C_Flat(params,
       base_optimizer, model, args)
```

>

- Boosting **8** SOTA methods
- Covering **all sorts of** CL

[NeurIPS 2024] "Make Continual Learning Stronger via C-Flat". 🐣 ♪: **C$_b$-a pitch that is one semitone lower than C tune.**

C-Flat (♪)
Framework

(a) Direct Tuning   (b) Regularization   (c) C-Flat

Legend:
- loss of old tasks
- loss of new tasks
- tolerance line
- perturb radius
- ☺ well learnt
- ☺ ordinary performance
- ☹ catastrophic forgetting

- **Continual Flatness with A Flatter Landscape**
  - ➢ Seeking flat minima that lie in neighborhood
  - ➢ Constraining the uniform curvature of the landscape
- **A Unified CL Framework using C-Flat**
  - ➢ Covering all sort of CL, Reg-based, Mem-based, Exp-based
  - ➢ Revisiting sharpness-aware minimization

4

[NeurIPS 2024] "Make Continual Learning Stronger via C-Flat". 🎏 ♪: C_b-a pitch that is one semitone lower than C tune.

C-Flat (♪)
Make Continual Learning Stronger

♪: **Plug-and-Play**

C-Flat: just **a line of code** suffices for its utilization.

```
1   from ... import C_Flat
2
3   # initialize optimizer for CL.
4   C_Flat_optimizer = C_Flat(params,
        base_optimizer, model, args)
```

➢ **Just a line of code**

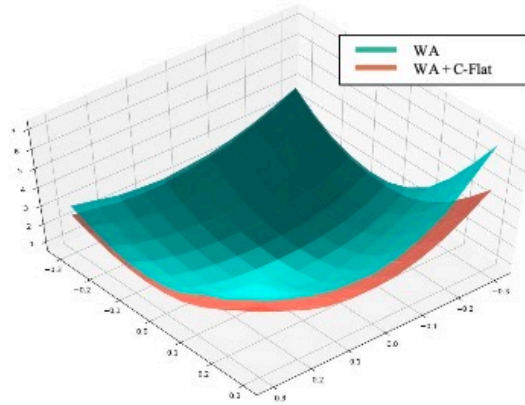| Method	Technology			CIFAR-100			ImageNet-100		Tiny-ImageNet
	Reg.	Mem.	Exp.	B0_Inc5	B0_Inc10	B0_Inc20	B50_Inc10	B50_Inc25	B0_Inc40
Replay [44]		•		58.83	58.87	62.82	**63.89**	72.18	43.31
w/ C-Flat				**59.98** ↑	**59.42** ↑	**64.71** ↑	63.60 ↓	**73.37** ↑	**44.95** ↑
iCaRL [43]		•		58.66	59.76	61.13	64.78	**77.25**	45.70
w/ C-Flat				**59.13** ↑	**60.40** ↑	**62.93** ↑	**65.01** ↑	76.22 ↓	**46.08** ↑
WA [64]	•			63.36	66.76	68.04	73.17	80.81	55.69
w/ C-Flat				**65.70** ↑	**67.79** ↑	**69.16** ↑	**73.56** ↑	**83.84** ↑	**56.06** ↑
PODNet [11]	•	•		48.05	56.01	63.45	83.66	85.95	54.24
w/ C-Flat				**49.70** ↑	**56.58** ↑	**64.37** ↑	**84.31** ↑	**86.85** ↑	**55.13** ↑
DER [57]			•	69.99	71.01	71.40	85.17	87.10	58.63
w/ C-Flat				**71.11** ↑	**72.08** ↑	**72.01** ↑	**86.64** ↑	**87.96** ↑	**60.14** ↑
FOSTER [54]	•		•	63.15	66.73	69.70	84.54	87.81	58.80
w/ C-Flat				**63.58** ↑	**67.34** ↑	**70.89** ↑	**85.40** ↑	87.81 -	**58.88** ↑
MEMO [68]			•	67.42	69.82	69.91	67.28	83.09	58.15
w/ C-Flat				**67.56** ↑	**69.94** ↑	**71.79** ↑	**69.34** ↑	**83.41** ↑	**58.97** ↑
Average Return				+1.04%	+0.66%	+1.34%	+0.62%	+0.90%	+0.81%
Maximum Return				+2.34%	+1.07%	+1.89%	+2.06%	+3.03%	+1.64%

Boosting **8** SOTA methods (**span all sorts of CL**)
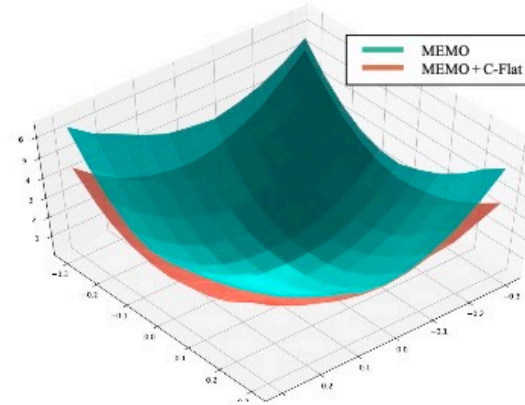
C-Flat (♪)
Visualization
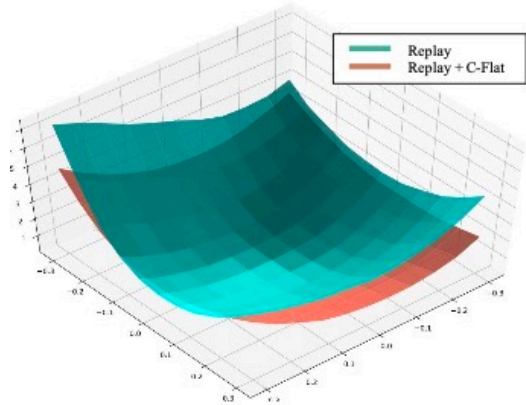
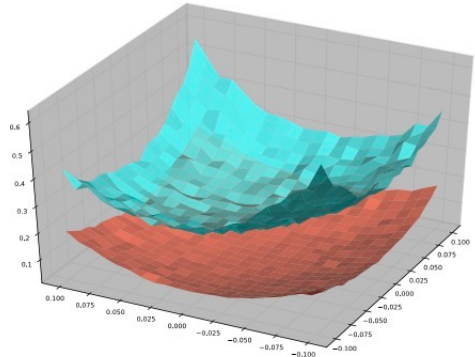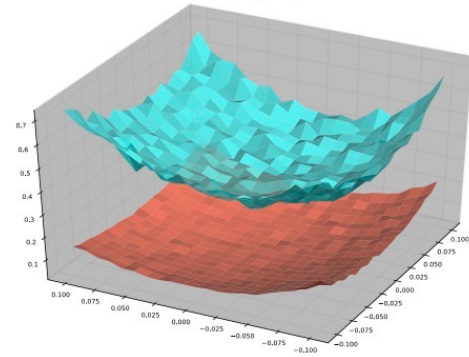(a) Replay w/ C-Flat   (b) WA w/ C-Flat   (c) MEMO w/ C-Flat)
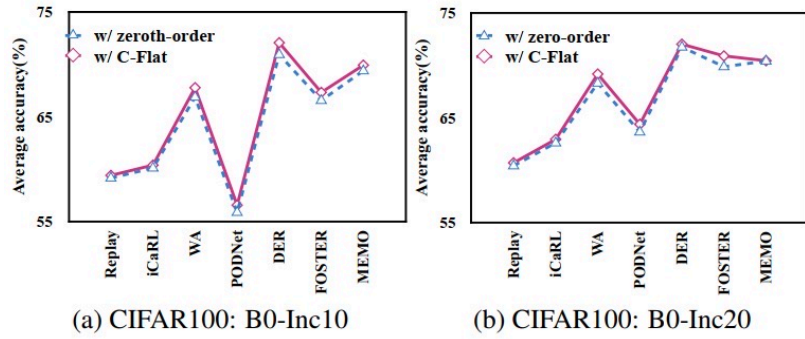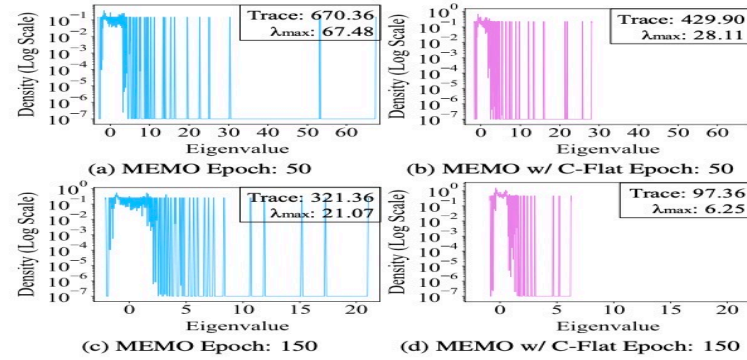
Global

Flatter is Better

(a) Replay   (b) WA   (c) MEMO

Local

[NeurIPS 2024] "Make Continual Learning Stronger via C-Flat". 🐣 ♪: C_b-a pitch that is one semitone lower than C tune.
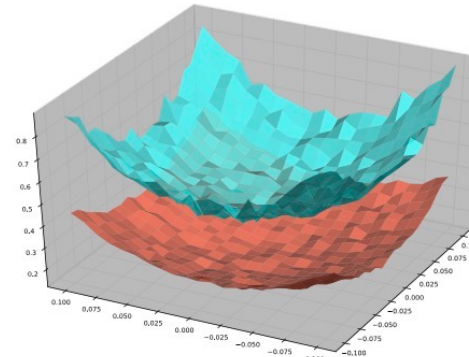
C-Flat (♪)
More discussions

Zeroth-order Flatness

(a) CIFAR100: B0-Inc10
(b) CIFAR100: B0-Inc20

Hessian Eigenvalues and Traces

(a) MEMO Epoch: 50 — Trace: 670.36, λmax: 67.48
(b) MEMO w/ C-Flat Epoch: 50 — Trace: 429.90, λmax: 28.11
(c) MEMO Epoch: 150 — Trace: 321.36, λmax: 21.07
(d) MEMO w/ C-Flat Epoch: 150 — Trace: 97.36, λmax: 6.25

Tier Guideline

Level	Speed	Boost (SGD/SAM)
L1	SGD > SAM > **C-Flat**	+2.39%/+1.91%
L2	SGD > **C-Flat** > SAM	+1.52%/+1.04%
L3	**C-Flat** > SGD	+1.51%/+1.03%

(a) Task 2 (b) Task 7 (c) Task 12 (d) Task 17

Across each task
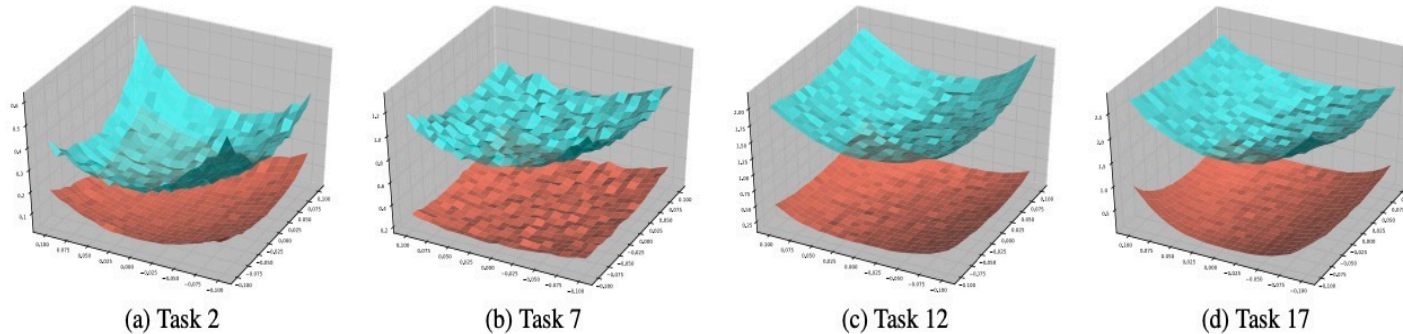
Beyond Not-Forgettting

Method		CIFAR-100/ B0_Inc5		
		w/o C-Flat	w/ C-Flat	RR
iCaRL [38]	old	36.36	37.12	BT+2.10%
	new	80.25	82.20	FT+2.43%
PODNet [9]	old	46.32	47.44	BT+2.42%
	new	62.65	64.75	FT+3.35%
FOSTER [48]	old	58.50	61.35	BT+2.85%
	new	62.05	63.05	FT+1.61%

[NeurIPS 2024] "Make Continual Learning Stronger via C-Flat". 🥚 ♪: **C_b-a pitch that is one semitone lower than C tune.**

Thank you for listening.