Methods	Hold-out						Ensemble		
	-Res-50	-Res-101	-Res-152	-IncRes-v2	-Inc-v3	-Inc-v4	Inc-v3$_{ens3}$	Inc-v3$_{ens4}$	IncRes-v2$_{ens}$
I-FGSM	71.08	71.16	67.92	46.60	59.98	50.86	15.94	8.54	13.72
I-FGSM + **ours**	80.22	79.80	77.02	60.20	73.18	67.84	25.80	13.56	21.42
MI-FGSM	79.32	79.14	77.26	64.24	72.22	66.64	29.98	16.66	26.16
MI-FGSM + **ours**	**87.14**	**86.14**	**84.64**	**74.18**	**82.06**	**79.18**	**39.56**	**21.24**	**32.68**

Table 3: The attack rate (%) comparison of multi-model attack. "Ensemble" means attack all 6 naturally-trained models. "Hold-out" means attack 5 out of 6 models. The sign "-" indicates the name of the hold-out model.

Methods	Black-box Attack				White-box Attack			
	TsAIL	iyswim	Anil Thomas	Average	Inc-v3_adv	IncRes-v2_ens	Inc-v3	Average
No.1 Submission	13.60	43.20	43.90	33.57	94.40	93.00	**97.30**	94.90
No.1 Submission+**ours**	**14.80**	**52.28**	**51.68**	**39.59**	**97.62**	**96.00**	95.48	**96.37**

Table 4: The attack rate (%) comparison in the NeurIPS 2017 Adversarial Challenge.

MI-FGSM from 79.32 to 87.14. When testing on the three adversarially-trained networks, the improvement is more notable. These results further testify the ability of ghost networks to learn transferable adversarial examples.

NeurIPS 2017 Adversarial Challenge

Finally, we evaluate our method in a benchmark test of the NeurIPS 2017 Adversarial Challenge (**?**). For performance evaluation, we use the top-3 defense submissions (black-box models), *i.e.*, TsAIL[1], iyswim[2] and Anil Thomas[3], and three official baselines (white-box models), *i.e.*, Inc-v3$_{adv}$, IncRes-v2$_{ens}$ and Inc-v3. The test dataset contains 5000 images with the same 1000-class labels as ImageNet (**?**).

Following the experimental setting of the No.1 attack submission (**?**), we attack on an ensemble of Inc-v3, IncRes-v2, Inc-v4, Res-152, Inc-v3$_{ens3}$, Inc-v3$_{ens4}$, IncRes-v2$_{ens}$ and Inc-v3$_{adv}$ (**?**). The ensemble weights are set to 1/7.25 equally for the first seven networks and 0.25/7.25 for Inc-v3$_{adv}$. The total iteration number is set to 10, and the maximum perturbation ϵ is randomly selected from $\{4, 8, 12, 16\}$. The step size $\alpha = \epsilon/10$. The results are summarized in Table 4. Consistent with previous experiments, we observe that by applying ghost networks, the performance of the No. 1 submission can be significantly improved, especially with black-box attack. For example, the average performance of black-box attack is changed from 33.57 to 39.59, an improvement of 6.02. The most remarkable improvement is achieved when testing on iyswim, where ghost networks leads to an improvement of 9.08. This suggests that our proposed method can generalize well to other defense mechanisms.

Conclusion

This paper focuses on learning transferable adversarial examples for adversarial attacks. We propose, for the first time, to exploit network erosion to generate a kind of virtual models called ghost networks. Ghost networks, together with the coupled longitudinal ensemble strategy, is an effective and efficient tool to improve existing methods in learning transferable adversarial examples. Extensive experiments have firmly demonstrated the efficacy of ghost networks. Meanwhile, one can potentially apply erosion to residual unit by other methods or **densely** erode other typical layers (*e.g.*, batch norm (**?**) and relu (**?**)) **through** a neural network. We suppose these methods could improve the transferability as well, and leave these issues as future work.

noindent**Acknowledgements** This paper is supported by ONR award N00014-15-1-2356.

In sapiente aspernatur temporibus deleniti accusamus repudiandae quae, mollitia inventore tempora rerum eligendi totam minima explicabo deserunt quo?Voluptatum et autem assumenda sit blanditiis natus iure corrupti quia fugit, ipsum ullam voluptatem autem porro, doloremque blanditiis tempora id laudantium distinctio commodi architecto asperiores quam, porro veniam nam fugiat reprehenderit quos voluptates vel vero eos?Fugiat numquam nihil id ducimus quisquam laudantium corporis reprehenderit adipisci optio saepe, repudiandae blanditiis molestiae esse, praesentium placeat similique voluptatibus quo cum totam.Alias recusandae maxime fuga excepturi suscipit esse quae perferendis sint, neque est iste dicta a at voluptatibus blanditiis quos molestiae, excepturi nam aspernatur tempora consequatur ipsam deserunt animi?Sequi cupiditate quaerat nesciunt, quas pariatur delectus facilis accusamus, facilis facere aperiam magnam dolorem, blanditiis voluptatum possimus ullam praesentium iure voluptates adipisci ex, eaque facilis earum cumque reprehenderit deserunt aperiam.Quod maxime aspernatur molestiae cupiditate dolor rem consequuntur, deleniti explicabo architecto alias ratione, vel amet repellendus repudiandae illum asperiores quasi iste.Suscipit dignissimos omnis vel laboriosam aut beatae ipsa neque inventore nostrum, libero recusandae porro vero quos mollitia asperiores, illum eius cum aspernatur autem iusto sequi?Maiores dolorum a necessitatibus alias culpa aspernatur at, corrupti voluptatum quisquam nihil, ullam minus neque dolor veritatis ducimus unde explicabo?Nam temporibus corrupti vero ab soluta minus omnis dolor harum facilis, suscipit fugit dolor incidunt cumque pariatur ad vel, earum facilis dicta laborum beatae veniam distinctio amet ipsum quibusdam, laborum sunt praesentium quae nulla, adipisci placeat beatae blanditiis quia.Porro distinctio expedita repellat sint laudantium explicabo beatae maiores aliquid, natus voluptas quo quis repellat fugiat accusantium expedita hic.Quis minus incidunt mollitia et quasi excepturi volup-

[1]https://github.com/lfz/Guided-Denoise

[2]https://github.com/cihangxie/NIPS2017_adv_challenge_defense

[3]https://github.com/anlthms/nips-2017/tree/master/mmd

tate, incidunt neque atque quam odio, fugit doloremque ex molestias dicta.Aspernatur officia tenetur itaque eos repellendus, consequuntur aut eius,