پنجمین کنفرانس ملی کامپیوتر ، فناوری اطلاعات و
کاربردهای هوش مصنوعی
۱۵ اسفندماه ۱۴۰۰

یادگیری نیمه نظارتی با استفاده از دانشجویان نویزدار

راشد اکبری، دانشجوی کارشناسی ارشد، akbari.r@qut.ac.ir [۱]
عبدالرضا رسولی کناری، دکتری و استادیار، rasouli@qut.ac.ir [۱]
مرتضی محجل کفشدوز، دکتری و استادیار، mohajjel@qut.ac.ir [۱]

چکیده

برای موفقیت یادگیری عمیق [۲]وجود منابع دادهای بزرگی که توسط متخصص و با صرف هزینه برچسب [۳]گذاری شده اند، ضروری هستند. اما زمانی که برچسب زنی دادگان، هزینه بر است، آموزش شبکه های عمیق با یک مجموعه کوچک برچسب دار، معمولا منجر به عملکرد قابل قبولی نمی شود. هدف از یادگیری نیمه نظارتی [۴]به کارگیری داده های بدون برچسب رهاشده ای است که به راحتی می توان آنها را گردآوری کرد. الگوریتم های نیمه نظارتی جدید که مبتنی بر افزون سازی دادها [۵]هستند توانسته اند به پیشرفت های جدیدی در این عرصه دست یابند. در این مقاله، با استفاده از الگوریتم دانشجویان نویزدار [۶] بر روی دادهای کم با استفاده از داده های بدون برچسب دقت مدل را بالا برده و با مشکلات ناشی از داده های کم فایق آید. روش پیشنهادی، مدل را تشویق می کند با آموزش دانشجویان مختلف خطای کلاس بندی را کاهش داده و با انجام این عمل به صورت متوالی، به واسطه خودآموزی بر روی داده هابدون برچسب عملکرد مدل تقویت می شود. روش پیشنهادی نتایج جدیدی در چهار مجموعه داده بدست آورده است. در حالی که در مجموعه داده های stl10 مدل تمام نظارتی با در اختیار داشتن ۵۰۰۰ نمونه برچسب دار، دقت ۶۳,۶ درصدی دارد، مدل پیشنهادی تنها با در اختیار داشتن ۵۰۰۰ نمونه برچسب دار و ۱۰۰۰۰۰ نمونه غیر برچسب دارتوانسته است به دقت ۹۴,۳ درصد برسد. آزمایشات مختلف مقیاس پذیری قدرت تعمیم روش پیشنهادی را چه در راستای افزایش اندازه مجموعه داده های برچسب دار و چه در راستای افزایش اندازه مدل نشان می دهد.

کلیدواژه ها: یادگیری نیمه نظارتی، یادگیری عمیق، دسته بندی، خودآموزی، افزون سازی داده

---

[۱] دانشگاه صنعتی قم

[2] Deep learning
[3] Label
[4] Semi-supervised learning
[5] Data Augmentation
[6] Noisy Students

۱- مقدمه

شبکه های عصبی عمیق در حال حاضر به صورت گسترده در بینایی ماشین و پردازش زبان های طبیعی مورد استفاده قرار میگیرند. موفقیت آنها تقریبا به واسطه مقیاس پذیری آنهاست، به این معنی که هر چه اندازه داده های آموزش بیشتر شود، عملکرد آنها نیز بهبود می یابد. شبکه های عصبی عمیق، برای رسیدن به چنین نتایجی معمولا نیاز به تعداد زیادی پارامتر دارند که آنها را مستعد بیش برازش شدن می کند. به همین دلیل، معمولا نیازمند یک مجموعه داده برچسب دار عظیمی هستند و عملکرد قوی خود را به واسطه یادگیری با نظارت بدست می آورند. اما جمع آوری داده هابرچسب دار برای بسیاری از مسائل هزینه بر است، چون نیاز به دانش یک متخصص دارد که آنها را برچسب بزند. امروزه بواسطه اینترنت در بسیاری از حوزه ها داده های خام بطور انبو در دسترس میباشد. به طور خلاصه، در بسیاری از مسائل، جمع آوری داده های بدون برچسب به مراتب کار آسان تر و کم هزینه تری است. یکی از راهکارهای موفق برای به کارگیری مجموعه بزرگی از داده های بدون برچسب، استفاده از یادگیری نیمه نظارتی است. از آنجا که داده های بدون برچسب، معمولا با کمترین نیاز به یک انسان متخصص بدست می آیند، هر بهبودی در عملکرد که از طریق یادگیری نیمه نظارتی حاصل شود، کم هزینه خواهد بود. به همین دلیل، تلاش های بسیاری برای توسعه الگوریتم ها و روشهای نیمه نظارتی صورت گرفته است. در این روش ها، معمولا تکنیکهایی ارائه می شود که با به کارگیری داده های بدون برچسب از بیش برازش شدن شبکه جلوگیری می کند. به این ترتیب، تعمیم پذیری مدل افزایش پیدا میکند[1].

آموزش شبکه های عصبی عمیق با استفاده از برچسب های محدود، در سال های اخیر مورد توجه قرار گرفته است. تفاوت روشهای مختلف یادگیری نیمه نظارتی در نحوه به کارگیری داده های بدون برچسب است. برخی از این روشها صرفا از داده های بدون برچسب برای یادگیری بدون نظارت[۷] استفاده میکنند. [2]. در دسته دیگر، سعی می شود به نحوی برچسب هایی برای نمونه های بدون برچسب تولید شود و از آنها در حین آموزش مدل استفاده کنند [3] . به این برچسب های تولید شده، شبه برچسب[۸] میگویند. این روشها از مشکل تأیید برچسب رنج می برند و پیش بینی اشتباه برای داده های بدون برچسب فرآیند آموزش مدل را دشوارتر می کنند[]. اخیرا با تزریق نویز به ورودی مدل بمنظور بالا بردن استحکام[۹] مدل در برابر تغییرات [4] موفقیت خوبی در حوضه یادگیری نیمه نظارتی عمیق بدست آمد. در این مقاله نشان داده شد روش های افزون زایی داده ها چقدر در بالا بردن استحکام مدل نقش مهمی دارند. ایده استحکام مدل با تزریق نویز در الگوریتم نیمه نظارتی خود یادگیری دانشجوی نویز دار [3] نیز بکار گرفته شد.

۱-۱- بیان مسئله

یادگیری نیمه نظارتی خود یادگیر، ترکیبی از یادگیری بانظارت و یادگیری بدون نظارت است. در این حالت، داده ها به دو مجموعه تقسیم می شوند: مجموعه برچسب دار به تعداد محدود (D_l) و مجموعه غیربرچسب دار به تعداد زیاد (D_u). روش های دسته بندی نیمه نظارتی تلاش می کنند، با به کارگیری D_u مدلی بدست آورند که عملکرد آن، در مقایسه با مدلی که تنها بر روی D_l آموزش دیده است، بهتر باشد .یک مشکل اساسی که در الگوریتم های خود یادگیر[۱۰] بروز میکند خطا در دسته بندی داده های غیر برچسب

[7] Unsupervised Learning
[8] Pseudo Label
[9] Robustness and Consistency
[10] Self-Training

دار می باشد. در الگوریتم دانشجوی نویزدار از مجموعه داده برچسب دار بزرگ ایتفاده کرده و هدف آن بالا بردن دقت بر روی کل مجموعه بوده است. بعلت اینکه از کل مجموعه برای اموزش استفاده شده خطای دسته بندی برا برچسب گذاری در حدود ۲٪ بوده است [3] که در مقابل کل مجموعه قابل چشم پوشی است ولی در مجموعه های برچسب دار محدود این خطا در حدود ۴۰٪ میباشد که رقم بزرگی است. اگر این مقدار داده برچسب دار اشتباه وارد مجموعه آموزشی شود قطعا یادگیری نیمه نظارتی نمی تواند بهبودی در مقایسه با یادگیری نظارتی داشته باشد. حتی ممکن است با گمراه کردن مدل، دقت پیش بینی را نیز کاهش دهد و نتیجه کلی مناسب نخواهد بود.

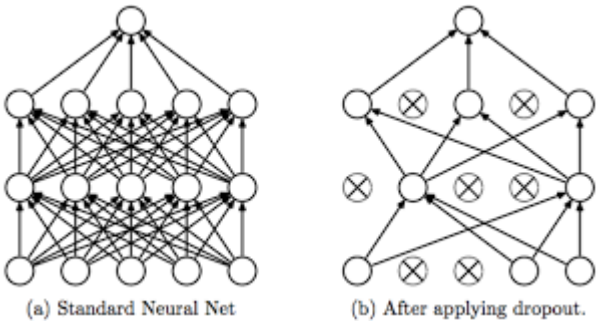
۲- شیوه انجام تحقیق

۲-۱- نویز

اصلی ترین بخش یادگیری نیمه نظارتی عمیق[11] تزریق نویز به مدل برای افزایش استحکام مدل است. در این مقاله از ۳ روش برای نویزدار کردن مدل دانشجو استفاده شده است.

۲-۱-۱- Drop Out

واژه‌ی dropout در لغت به معنای بیرون رفتن یا رها کردن است. در هنگام آموزش، برخی از سلول های عصبی[12] به صورت تصادفی خروجی صفر می‌گیرند(شکل ۱). این بدین معنا است که آموزش برروی معماری‌های مختلف با مجموعه سلول های عصبی متفاوت انجام می‌گیرد. می‌توانید به dropout به عنوان یک تکنیک گروهی نگاه کنید که در آن خروجی چندین شبکه مخلوط می‌شوند تا خروجی نهایی را شکل دهند[5].

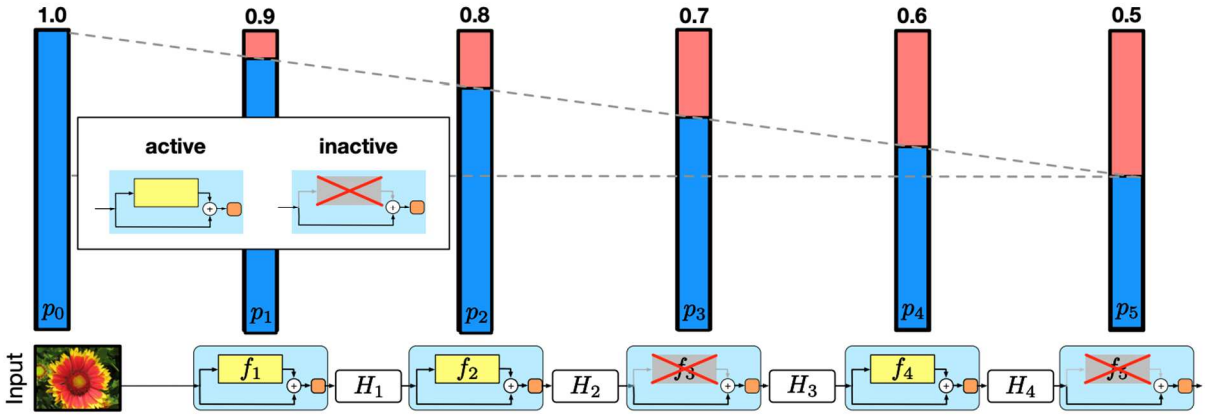


شکل ۱ DropOut

[11] Deep Semi-Supervised Learning
[12] neuron cell

۲-۱-۲- عمق تصادفی[13]

عمق تصادفی همانند DropOut در سطح لایه ها عمل میکند. در هر دور آموزش بطور تصادفی تعدادی از لایه ها حذف میشوند (شکل ۲) و بدین وسیله ۲ هدف دنبال میشود. هدف اول کاهش طول شبکه عصبی و در نتیجه کاهش زمان آموزش و هدف دوم جلوگیری از بیش برازش[14] . در این مقاله از عمق تصادفی برای بالا بردن استحکام مدل بعنوان تولید کننده نویز در سطح لایه ها استفاده شده است[5].



شکل ۲ عمق تصادفی

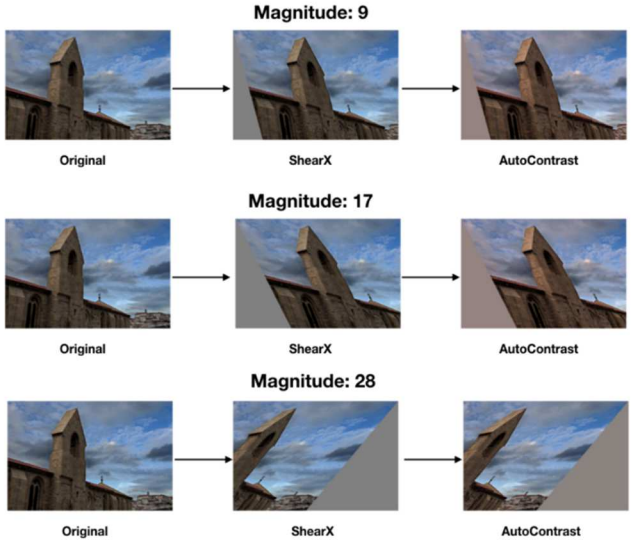
۲-۱-۳- افزون سازی داده

مهمترین و اصلی ترین روش افزودن نویز به مدل دانشجو در این مقاله افزون سازی داده بصورت شدید است. مدل های دانشجو در زمان آموزش با شدیدترین نوع تغییرات بر روی داده های ورودی مواجه میشوند تا استحکام مدل در بالاترین حد ممکن قرار گیرد. در این مقاله از روش افزون سازی داده RandAugment [7] (شکل ۳) استفاده شده که با کاهش فضای جستجو و ارایه اخرین تکنیک های افزون سازی داده سرعت و کیفیت آموزش را بالاتر برده است.

---

[13] Stochastic Depth

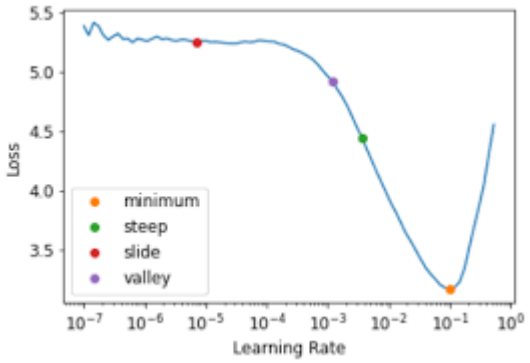
[14] Overfit



شکل ۳ RandAugmentation

۲-۲- پارامترهای بهینه و جلوگیری از بیش برازش

در روش‌های معمول پیدا کردن نرخ یادگیری که یکی از هایپرپارامترهای مهم در یادگیری می باشد اغلب از کاهش نرخ یادگیری استفاده می شد در این روش غیر از کم کردن گاهی نرخ یادگیری افزایش می دهند به عبارتی نرخ یادگیری افزایشی و کاهشی باشد تا در صورت افتادن در کمینه محلی بتواند از آن خارج شود در واقع کار اصلی آن پیدا کردن نرخ یادگیری مناسب می باشد. اگر نرخ یادگیر را اشتباه بدهیم مدل واگرا شده و به جواب مورد نظر نمی رسد. ابتدا با یک نرخ یادگیری کوچک مدل شروع به یادگیری می کند برای جلوگیری از کند شدن فرآیند یادگیری به جای هر کل داده، بخش کوچکی از داده به مدل داده میشود و بعد از بار آموزش مقدار خطا (Loss) ذخیره و نرخ یادگیری با یک مقدار مشخص شده افزایش پیدا می کند. در نهایت با محاسبه خطای کمینه، نرخ یادگیری متناسب با ان را برمیگرداند (شکل ۴).



شکل ۴ نرخ یادگیری بهینه

از CAllBack ها برای جلوگیری از بیش برازش در حین آموزش استفاده شده است. اولین آنها EarlyStopping است که برای تحت نظر گرفتن validation loss است. در صورتی که درحین آموزش مقدار آن کم نشود آموزش را متوقف میکند. دیگیری ذخیره بهترین مدل است یعنی ما از اینکه نتیجه آموزش یک مدل با بهترین وزن هاست که این callback این وظیفه را انجام میدهد.

۳- روش دانشجویان نویزدار

۳-۱- آموزش مدل استاد

ابتدا مدل استاد را با استفاده از داده‌های برچسب دار آموزش می‌دهیم و با استفاده از مدل آموزش دیده داده های غیربرچسب دار را برچسب گذاری میکنیم.

۳-۲- حذف داده های غیرمرتبط

داده های غیر برچسب دار ممکن است حاوی سمپل هایی باشند که به هیچ یک از کلاس های ما تعلق نداشته باشند. برای جداسازی آنها از حد[۱۵] پایین نتیجه پیش بینی مدل استاد استفاده میکنیم. برای اینکار بیشترین مقدار پیش بینی به کلاس ها را با مقدار حد پایین مقایسه میکنیم و در صورتی که کمتر از آن بود حذف شده و در غیر اینصورت به مرحله بعد فرستاده میشود.

۳-۳- متعادل کردن کلاس ها

در اکثر موارد خروجی مرحله قبل یک مجموعه داده غیرمتعادل است. یکی از چالش های یادگیری عمیق چالش مجموعه داده های غیرمتعادل است. در این مقاله از ۲ روش برای متعادل سازی کلاس ها استفاده شده است. در ابتدا تعداد داده های هر کلاس محاسبه شده و بیشترین مقدار داده بعنوان هدف در نظر گرفته میشود. مقادیر مورد نیاز برا هر کلاس با تفاضل ماکزیمم از مقدار هر کلاس محاسبه شده و به دو روش افزایش داده و تولید داده های مصنوعی توسط شبکه های عصبی مولد متخاصم[۱۶] بصورت مساوی اضافه میشود. باید در نظر داشته باشیم که ورودی هر دو روش داده برچسب دار میباشند نه برچسب خورده ها زیرا هنوز احتمال وجود داده های برچسب خورده اشتباه وجود دارد که با این روش ها مقدار خطا را افزایش خواهیم داد که نتیجه کلی گمراه کننده خواهد بود. روش افزایش داده همان RandAugment میباشد و از Conditional GAN[۱۷] [8] برای تولید داده های جدید استفاده شده است.

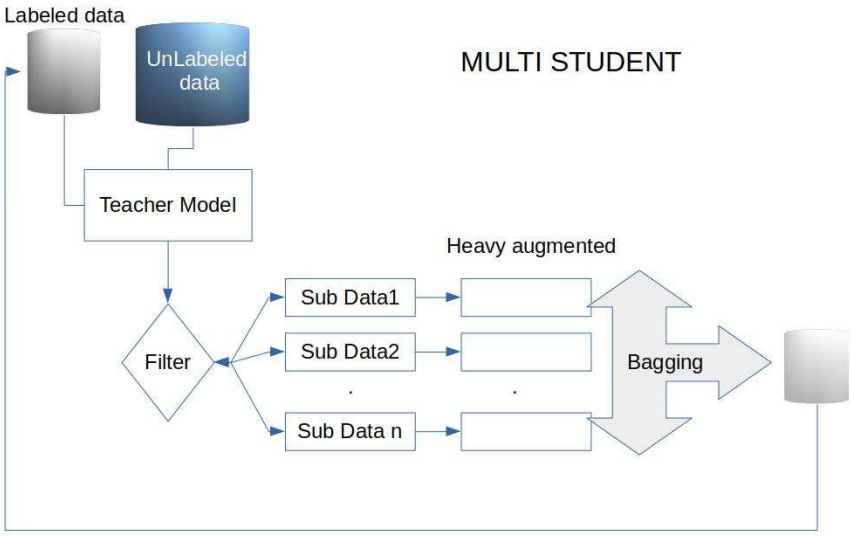
[15] Threshold
[16] Generative Adversarial Network
[17] CGAN

۳-۴- تقسیم داده ها

داده های حاصل شده از مرحله قبل به n قسمت تقسیم میشود. از بین آنها n-1 بخش را انتخاب کرده و بتوسط آنها n-1 دانشجو را آموزش داده میشود. برای آموزش دانشجوها از نویز که در بخش ۲-۱ توضیح داده شد بصورت شدید استفاده میکنیم. یکی از دلایل بهتر عمل نمودن دانشجو ها از استاد بزگتر بودن مدل و داده های آموزشی دانشجویان حتی ئس از تقسیم داده هاست.

۳-۵- رای گیری

بخش n داده ها که هیچ کدام از دانشجویان با آن آموزش ندیده اند به تمام دانشجویان تزریق میکنیم تا بر روی آن پیش بینی انجام دهند. بر روی n-1 پیش بینی رای گیری[18] کرده و در صورتی که نتیجه رای گیری فقط یک کلاس بود داده را به داده های برچسب دار اضافه کرده و از بین داده ها حذف میکنیم. بخش ۳ را به دفعات تا رسیدن به نتیجه دلخواه تکرار میکنیم که در شکل ۵ شمای کلی آن را میبینیم.



شکل ۵ Multi Student

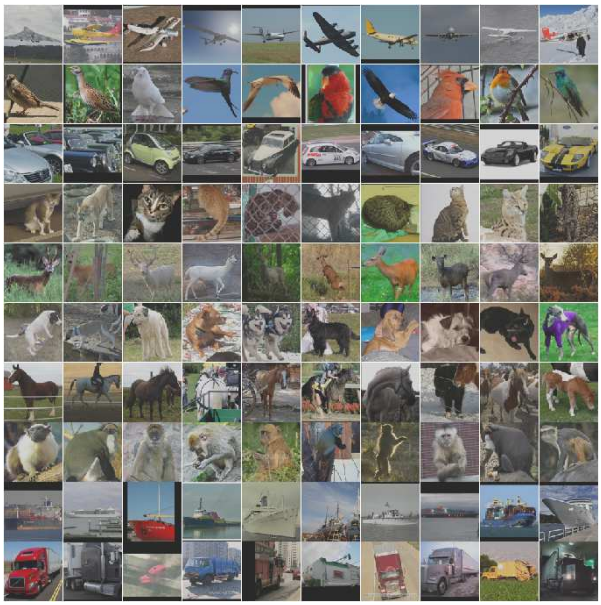
[18] Bagging

۴- نتایج

۴-۱- مجموعه داده ها

۴-۱-۱- Stl10

مجموعه داده stl10 [9] از دو بخش برچسب دار و غیر برچسب دار تشکیل شده است که مخصوصا برای آموزش های نمیه نظارتی مورد استفاده قرار میگیرد. تمامی عکس های این مجموعه ۹۶ در ۹۶ پیکسل تشکیل شده و توسط دانشگاه استنفورد جمع آوری شده است. بخش اول دارای ۱۰ کلاس و شامل ۵۰۰ عدد برای هر کلاس بصورت رنگی میباشد. بخش دوم آن شامل ۱۰۰۰۰۰ عکس رنگی غیربرچسب دار است (شکل ۶).



شکل ۶-STL10

۴-۲- CIFAR10

این مجموعه داده شامل ۶۰۰۰۰ عکس به اندازه (۳۲در۳۲) بصورت رنگی در ۱۰ کلاس می باشد. این مجموعه توسط دانشگاه تورنتو تهیه شده است. معمولا برای یادگیری نظارتی مورد استفاده قرار میگیرد ولی در این مقاله به دو بخش ۵۰۰۰ تایی برچسب دار و ۵۵۰۰۰ تا غیر برچسب دار تقسیم شده تا ارزیابی بر روی تعداد خطاهای دسته بندی انجام شود[10].

۵- نتایج آزمایشات

نتایج آزمایشات در جدول ۱ نشان داده شده است.

جدول ۱- خلاصه نتایج آزمایش ها.

Data Set Name	Hyper Parameters	Accuracy on D_l	Final Accuracy
Stl10	batch_size=128, learning_rate=0.001 pooling = max_pooling	63.6	**92.2**
Stl10	batch_size=64, learning_rate=0.003 pooling = avg_pooling	76.5	**94.3**
Cifar10	batch_size=128, learning_rate=0.003 pooling = max_pooling	66.6	**95.3**
Cifar10	batch_size=64, learning_rate=0.001 pooling = avg_pooling	78.4	**96.7**

۶- نتیجه گیری

روش دانشجوی نویز دار فقط بر روی مجموعه دادهای بزرگ ازمایش شده بود در این مقاله هدف ما ارایه روش دانشجویان نویز دار و آزمایش آن بر روی مجموعه داده های کوچک و محدود بود. نتایج حاصله که در جدول ۱ قابل مشاهده است نشان از افزایش تقریبا ۳۰٪ بر روی هر دو مجموعه داده است. با وجود اینکه روش پیشنهادی زمان اجرای طولانی تری نسبت به روشهای یادگیری با نظارت دارد ولی نتایج بدست آمده نشان میدهد که ارزش کار و وقت صرف شده را دارد. بنابر این در حالتی که داده های فراوان در دسترس باشد براحتی میتوان از این روش برای بالا بردن دقت مدل استفاده کرد

مراجع

[1] A. Oliver, A. Odena, C. Raffel, E. D. Cubuk, and I. J. Goodfellow, “Realistic evaluation of deep semi-supervised learning algorithms,” in NeurIPS, 2018, pp. 3239–3250.

[2] J. T. Springenberg, “Unsupervised and semi-supervised learning with categorical generative adversarial networks,” in ICLR, 2016.

[3] Q. Xie, M. Luong, E. H. Hovy, and Q. V. Le, “Self-training with noisy student improves imagenet classification,” in CVPR. IEEE, 2020, pp. 10 684–10 695.

[4] Q. Xie, Z. Dai, E. H. Hovy, T. Luong, and Q. Le, “Unsupervised data augmentation for consistency training,” in NeurIPS, 2020.

[5] S. Park, J. Park, S. Shin, and I. Moon, “Adversarial dropout for supervised and semi-supervised learning,” in AAAI. AAAI Press, 2018, pp. 3917–3924

[6] G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Weinberger, “Deep Networks with Stochastic Depth,” *arXiv:1603.09382 [cs]*, Jul. 2016, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1603.09382

[7] E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le, “RandAugment: Practical automated data augmentation with a reduced search space,” *ArXiv190913719 Cs*, Nov. 2019, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1909.13719

[8] M. Mirza and S. Osindero, “Conditional Generative Adversarial Nets,” *ArXiv14111784 Cs Stat*, Nov. 2014, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1411.1784

[9] “STL-10 dataset.” https://cs.stanford.edu/~acoates/stl10/ (accessed Feb. 11, 2022).

[10] “CIFAR-10 and CIFAR-100 datasets.” https://www.cs.toronto.edu/~kriz/cifar.html (accessed Feb. 11, 2022).