Samonadgledano učenje pomoću pseudo-zadatka rotacije na posebno dizajniranoj arhitekturi duboke neuronske mreže

Milutin Cerović

***Sadržaj* — Za obučavanje dubokih neuronskih mreža i postizanje dobrih performansi u slučajevima učenja vizuelnih karakteristika iz slika ili videa za aplikacije u računarskoj viziji neophodne su velike količine označenih podataka. Da bi se izbegao trošak sakupljanja i označavanja velikih skupova podataka, kao podskup metoda nenadgledanog učenja, ističu se metode samonadgledanog učenja za rešavanje ovog problema. One uspevaju da nauče opšte vizuelne karakteristike slika i videa iz neoznačenih skupova podataka. U radu je implementirana konvoluciona neuronska mreža koja ima pseudo-zadatak da prepozna koja geometrijska transformacija je primenjena za sliku sa ulaza, to jest da prepozna pod kojim uglom je slika rotirana. Nakon obuke tehnikama transfera znanja na malom podskupu označenih podataka mreža je obučavana za zadatak klasifikacije slika. Na referentnom skupu podataka, STL10, postignuta je tačnost od 76% na ciljnom zadatku klasifikacije slika.**

***Ključne reči* — ciljni zadatak, kontrastivno učenje, neoznačeni podaci, pseudo-zadatak, računarska vizija, samonadgledano učenje, vizuelne karakteristike**

I. Uvod

U današnjem svetu postoji enormna količina podataka, koja svakog dana nezaustavljivo raste, koju je teško sistematizovati i objasniti značenje svakog podatka i šta on predstavlja. Algoritmi nenadgledanog mašinskog učenja imaju mogućnost da samostalno nauče karakteristike datog skupa podataka bez pomoći eksperta i eksplicitnog označavanja (labeliranja) podataka. U slučaju

Milutin Cerović, Računarski fakultet Union Univerzitet, Knez Mihailova 6/6, 11000 Beograd, Srbija; (e-mail: milutin.cerovic@gmail.com).

kada je dato dovoljno označenih podataka, govorimo o algoritmima nadgledanog mašinskog učenja koji mogu da reše probleme vrlo uspešno i značajno efikasnije od nenadgledanog mašinskog učenja. Međutim, dobre performanse iziskuju veću količinu označenih podatka. Manuelno prikupljanje i označavanje podataka, kao kod npr. ImageNet [1] skupa podataka, je vremenski skupo i kompleksno za skaliranje. Razmatrajući količinu neoznačenih podataka (npr. tekst ili slike na Internetu) jasno je da je ona značajno veća od ograničenog skupa podataka koji ljudi kreiraju. Postavlja se jasno pitanje, kako iskoristiti ogromnu količinu dostupnih podataka, bez eksplicitnog označavanja podataka za željeni zadatak?

Posebna paradigma mašinskog učenja pod nazivom samonadgledano učenje omogućava da se iskoriste ogromne količine neoznačenih podataka i to na dva moguća načina: konstruišući pseudo-zadatak nadgledanog učenja koji se obučava nad neoznačenim podacima ili pomoću kontrastivnog učenja. U prvom pristupu zadatak koji se konstruiše kod samonadgledanog učenja navodi na funkciju troška nadgledanog učenja, ali u ovom slučaju nije fokus na performansama konstruisanog zadataka već na učenju neposrednih karakteristika sa očekivanjem da te karakteristike nose dovoljno semantičkog ili strukturalnog značenja i da mogu biti od benefita za ciljnih zadatak. Ovako naučene karakteristike je moguće iskoristiti naknadno uz pomoć transfera znanja na vrlo malom skupu označenih podataka za ciljni zadatak koji se rešava, npr. klasifikacija slika. Drugi pristup je zasnovan direktno na metrici udaljenosti koja ima za cilj da uspostavi sličnost ili različitost između podataka (npr. slika), čime se neposredno forsira da nauči njihova obeležja i karakteristike.

Postoje razne tehnike i metodologije kreiranja pseudo-zadataka, međutim svojom jednostavnošću kao i kvalitetom naučenih vizuelnih karakteristika ističu se pseudo-zadaci koji imaju za cilj da prepoznaju koja geometrijska transformacija je primenjena nad datim ulazom, specifično u ovom radu - geometrijska transformacija rotacije. Cilj rada je da predstavi značaj i mogućnosti ove metodologije - da je moguće naučiti opšte vizuelne karakteristike iz neoznačenog skupa podataka prepoznavanjem rotacije na dizajniranoj arhitekturi konvolucione mreže sa malim brojem slojeva i parametara. Ovakve arhitekture mreža su često neophodne kada se izvršavaju na ograničenim hardverskim resursima.

II. SAMONADGLEDANO UČENJE – PSEUDO ZADATAK

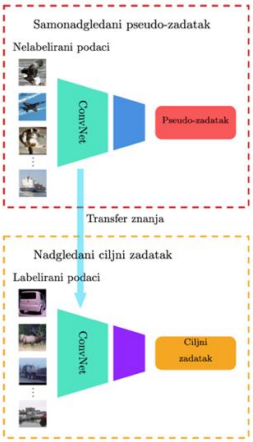
Ideja je da se na osnovu pseudo-zadatka automatski definišu pseudo oznake koje će se koristi u funkciji troška nadgledanog učenja. Dok mreža rešava dati

pseudo-zadatak na pseudo-oznakama ona u tom procesu obučavanja uči semantičke i strukturalne opšte vizuelne karakteristike. Tako naučena mreža može se iskoristiti za rešavanje ciljnog zadatka [2].

Za dati skup podataka $\mathcal{D} = \{(x_i, \mathrm{p_i}) \mid x_i \in \mathcal{X}, \mathrm{p_i} \in \mathrm{P}\}_{\mathrm{i=1}}^{\mathrm{M}}$, pri čemu je svakoj instanci x_i pridružena automatski generisana pseudo-labela p_i, funkcija troška se može definisati kao i u slučaju nadgledanog učenja:

$$Loss(\mathcal{D}) = \min_{\theta} \frac{1}{M} \sum_{1}^{\mathrm{M}} \mathrm{loss}(\mathrm{x_i}, \mathrm{p_i}) \qquad (1)$$

Generalni način rada samonadgledanog učenja je prikazan na Sl. 1. Tokom faze obuke samonadgledanog učenja, predefinisani pseudo-zadatak je dizajniran da konvolucioni modeli mogu da ga reše, dok su pseudo-oznake automatski generisane na osnovu nekih atributa podataka. Konvolucioni model se obučava da nauči ciljnu funkciju pseudo-zadatka (1). Nakon samonadgledane obuke modela, naučene vizuelne karakteristike mogu se dalje prebaciti za ciljne zadatke (posebno kada je dostupan relativno mali skup podataka). Generalno, početni slojevi uspevaju da nauče opšte karakteristike niskog nivoa kao što su ivice, uglovi i teksture, dok dalji slojevi uče karakteristike vezane za dati zadatak. Zato se karakteristike iz početnih slojeva koriste u transferu za nadgledani ciljni zadatak. Ciljni zadaci u kompjuterskoj viziji se koriste za evaluaciju kvaliteta karakteristika naučenih tokom samonadgledanog učenja.

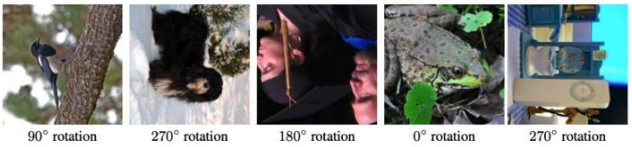


Sl. 1. Generalni prikaz rada samonadgledanog učenja u slučaju pseudo-zadatka.

III. PSEUDO-ZADATAK ROTACIJA

Inicijalna ideja predložena u radu [3] je da se učenje vizuelnih karakteristika slika u slučaju samonadgledanog učenja postigne obučavanjem konvolucionih mreža koje prepoznaju koja geometrijska transformacija je primenjena za sliku sa ulaza. Očekivano je da male distorzije na slici ne utiču na originalno semantičko značenje ili geometrijske forme. Ovako kreirane slike sa distorzijom se smatraju iste kao i originali, pa iz toga sledi da je učenje opštih vizuelnih karakteristika invarijantno na distorziju.

Autori su predložili diskretan skup geometrijskih transformacija koje se primenjuju na slici iz skupa za obuku. Tako transformisana slika prosleđuje se dubokoj konvolucionoj mreži koja se obučava da prepoznaje koja transformacija je primenjena na svakoj slici. Ovako formulisan, skup geometrijskih transformacija predstavlja pseudo-zadatak koji konvolucioni model treba da nauči. Da bi se postiglo nenadgledano učenje semantičkih karakteristika autori su predložili rotaciju slika za $0°, 90°, 180°$ i $270°$, Sl. 2. Konvolucioni model treba da prepozna koja od ovih rotacija je primenjena na sliku sa ulaza, što inherentno znači da model mora da razume osnovne koncepte objekata sa slike, kao što su lokacija, tip, poza, tekstura, ivice, itd.



Sl. 2. Prikaz rotacija iz skupa za obuku

Neka je definisan skup K diskretnih geometrijskih transformacija $G = \{g(\cdot\,|y)\}_{y=1}^{K}$ pri čemu je $g(\cdot\,|y)$ operator koji predstavlja geometrijsku transformaciju sa oznakom y primenjenu na sliku x, čime se dobija transformisana slika $x^y = g(x|y)$. Geometrijske transformacije G treba da definišu klasifikacioni zadatak koji implicitno forsira konvolucioni model da nauči semantičke karakteristike za zadatke vizuelne percepcije (npr. detekcija objekata ili klasifikacija slika). Predložen je konačan skup mogućih geometrijskih transformacija G koji obuhvata sve moguće rotacije slike sa umnoškom od 90 stepeni, npr. za 2d sliku te rotacije su 0, 90, 180 i 270 stepeni. Konvolucioni model $F(\cdot)$ kao ulaz dobija sliku x^{y*}, gde je oznaka y^* nepoznata modelu i daje izlaz koji predstavlja raspodelu verovatnoće za sve moguće geometrijske transformacije:

$$F(x^{y*}|\theta) = \{F^{y}(x^{y*}|\theta)\}_{y=1}^{K}, \quad (2)$$

Za dati skup podataka $\mathcal{D} = \{x_i\}_{i=0}^{M}$ cilj samonadgledanog obučavanja koje konvolucioni model treba da nauči da reši je:

$$\min_{\theta} \frac{1}{M} \sum_{i=1}^{M} \left(-\frac{1}{K} \sum_{i=1}^{K} \log\left(F^{y}(g(x_i|y)|\theta)\right) \right) \quad (3)$$

IV. Implementacija

Arhitektura modela, prikazana na Sl. 3. pravljena je po uzoru na jednostavnije konvolucione modele i srazmerna je u odnosu na dati problem i veličinu skupova podataka na kojima je obučavana, za razliku od većih i opštijih modela, kao što su ResNet [4], Inception [5] i VGG [6]. Problem navedenih arhitektura je što su one često nepotrebno velike i zahtevne prilikom izvršavanja, stoga je ideja da se iskoriste male arhitekture koje su efikasne kako za obuku tako i prilikom izvršavanja u aplikacijama.

U slučaju ciljnog zadatka početni deo arhitekture, *feature extractor*, ostaje nepromenjen i ne obučava se, već se dodaje nova glava mreže, Sl. 4. na neki od prethodnih konvolucionih blokova i jedino se ona obučava na nadgledan način pomoću označenog dela skupa podataka.

Optimizator korišćen prilikom obuke je Adam, [7], sa organizatorom stope učenja koja je inicijalno je postavljena na vrednost $1e^{-3}$ do 100 epohe obuke, a nakon toga na $1e^{-4}$. Aktivaciona funkcija u konvolucionim slojevima je ReLU [8]. Takođe su korišćeni slojevi za unutrašnju standardizaciju [9] i slojevi nasumičnog odbacivanja neurona za prevenciju *overfitting-a* [10].

Metode augmentacije podataka su takođe korišćene: transformacije prostora boja - svođenje na sivu sliku, kao i promene vrednosti kontrasta, osvetljenja i zasićenja boja.

Za implementaciju je korišćena visoko apstrahovana biblioteka za kreiranje i obuku modela dubokog učenja Keras [11] koja u pozadini koristi Tensorflow [12] biblioteku koja predstavlja jezgro i koja izvršava numerička izračunavanja korišćenjem paradigme protoka podataka i diferencijalnog programiranja.

Obuka je izvršavana na grafičkoj karti Nvidia GeForce RTX 2080 sa prosečnim trajanjem obuke približno 2 sata.

	Type	Number of filters	Kernel	Strides	Padding	Output Shape
Feature extractor	Convolution2D	32	3	1	1	32 × 32 × 32
	BatchNorm	-	-	-	-	32 × 32 × 32
	Convolution2D	32	3	1	1	32 × 32 × 32
	BatchNorm	-	-	-	-	32 × 32 × 32
	MaxPooling	1	2	2	-	16 × 16 × 32
	DropOut	-	-	-	-	16 × 16 × 32
	Convolution2D	64	3	1	1	16 × 16 × 64
	BatchNorm	-	-	-	-	16 × 16 × 64
	Convolution2D	64	3	1	1	16 × 16 × 64
	BatchNorm	-	-	-	-	16 × 16 × 64
	MaxPooling	1	2	2	-	8 × 8 × 64
	DropOut	-	-	-	-	8 × 8 × 64
	Convolution2D	128	3	1	1	8 × 8 × 128
	BatchNorm	-	-	-	-	8 × 8 × 128
	Convolution2D	128	3	1	1	8 × 8 × 128
	BatchNorm	-	-	-	-	8 × 8 × 128
	MaxPooling	1	2	2	-	4 × 4 × 128
	DropOut	-	-	-	-	4 × 4 × 128
	Flatten	-	-	-	-	2048
Head	Dense	-	-	-	-	128
	BatchNorm	-	-	-	-	128
	DropOut	-	-	-	-	128
	Dense	-	-	-	-	4

Sl. 3. Arhitektura modela. Svaki red u tabeli predstavlja jedan sloj u mreži, sa datim odgovarajućim brojem filtara. Takođe su dati veličina i način pomeranja filtara, i veličina ivica popunjenih nulama, kao i izlazni tenzor nakon trenutnog sloja.

	Type	Number of filters	Kernel	Strides	Padding	Output Shape
Head	Dense	-	-	-	-	128
	DropOut	-	-	-	-	128
	Dense	-	-	-	-	10

Sl. 4. Prikaz nove glave modela koja se obučava za potrebe ciljnog zadatka. Jedino se ažuriraju parametri glave, dok vrednosti parametara u ostatku mreže ostaju nepromenjeni

V. REZULTATI

Modeli mašinskog učenja se najčešće obučavaju i evaluiraju na standardizovanim skupovima podataka koji omogućavaju određenu sistematizaciju i mogućnost poređenja sa ostalim modelima. Metode

samonadgledanog učenja se mogu obučavati sa slikama odbacivanjem oznaka kreiranim od strane ljudi, stoga, svi skupovi podataka koji su prikupljeni za nadgledano učenje mogu se koristiti za učenje vizuelnih karakteristika. U cilju ovog rada korišćeni su skup STL10 [13]. Sastoji se od 5000 označenih slika za nadgledano obučavanje, 8000 za testiranje i 100000 neoznačenih slika. Slike za nadgledano obučavanje raspoređene su u 10 kategorija i fiksne su veličine, 96×96 piksela.

TABELA 1: PRIKAZ METRIKA ZA PSEUDO-ZADATAK

Ugao (klasa)	**Preciznost**	**Odziv**	**Broj uzoraka**
0	0.94	0.94	24894
90	0.94	0.95	24894
180	0.95	0.94	24894
270	0.94	0.94	25110
Makro-prosek	0.94	0.95	

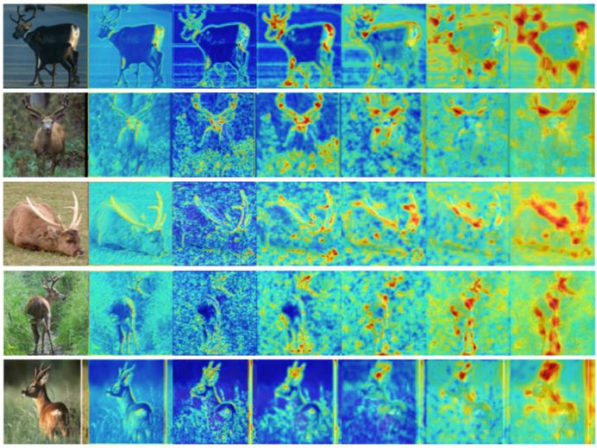
Može se zaključiti da model po svim metrikama ima veoma visoke vrednosti, što implicira da je savladao dosta dobro pseudo-zadatak. U nastavku je analiziran kvalitet tih karakteristika, kao i koliko je dobro generalizovao te karakteristike na ciljnom zadatku.

Vizuelnom inspekcijom utvrđeno je da model najčešće ima problema sa slikama koji su rotirane sa uglovima od 180° i 90°, jer u tim situacijama vrlo često nema semantičke greške i teško je doneti odluku, jer slika može biti orijentisana i na taj način što potvrđuju i primeri u skupu za obučavanje sa takvom orijentacijom, Sl. 5.

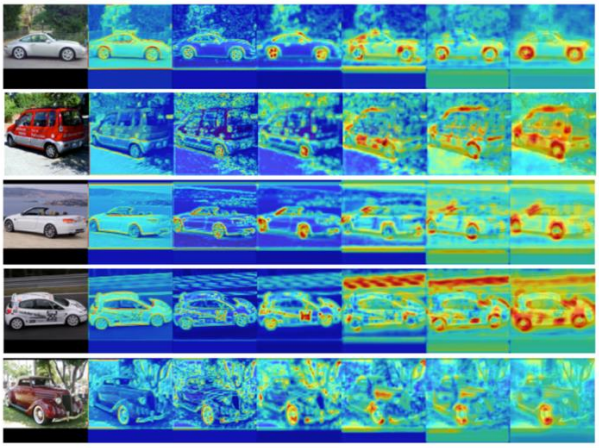
Sl. 5. Prikaz instanci na kojima model greši

Vizuelizacija aktivacija je veoma korisna za razumevanje toga kako uzastopni slojevi konvolucione mreže transformišu ulaze, kao i za dobijanje ideje o značenju pojedinačnih filtara. Na ovaj način može se videti dekompozicija slike kroz različite slojeve, kao i to na kojim delovima slike na svakom sloju mreža daje značaj pri donošenju odluka. Vizuelizacija je predstavljena kroz mapu aktivacija koja označava delove na koje mreža obraća najviše pažnje [14]. Na Sl. 6. se može videti da mreža tokom pseudo-zadatka zaista uči najznačajnije karakteristike određene klase. U slučaju klase „Jelen“

Sl. 6. (a) mreža je naučila da se fokusira na rogove, dok je u slučaju klase „Automobil“ fokus stavila na točkove Sl. 6. (b). Ovo ne znači da mreža donosi odluku samo na osnovu označenih karakteristika, ali one svakako predstavljaju nešto jedinstveno i specifično za datu klasu.



(a) Mapa aktivacija za instance klase „Jelen”



(b) Mapa aktivacija za instance klase „Automobil”

Sl. 6. Redovi predstavljaju različite instance iste klase iz skupa podataka, dok su kolone mape aktivacija sa početnih konvolicionih i agregacionih slojeva

Ovako naučene reprezentacije iskorišćene su za transfer znanja modela nadgledanog učenja korišćenjem nove glave modela na označenim podacima. Za novu mrežu i zadatak klasifikacije slika rezultati su dati u Tabela 2.

TABELA 2: PRIKAZ METRIKA NA CILJNOM ZADATKU

Klasa	**Preciznost**	**Odziv**	**Broj uzoraka**
Avion	0.88	0.88	800
Ptica	0.74	0.77	800
Automobil	0.93	0.87	800
Mačka	0.63	0.62	800
Jelen	0.68	0.72	800
Pas	0.61	0.50	800
Konj	0.75	0.81	800
Majmun	0.69	0.68	800
Brod	0.88	0.89	800
Kamion	0.82	0.88	800
Makro-prosek	0.76	0.76	

Makro-prosek za preciznost i odziv iznosi 76% i 76%, respektivno. Dobijena tačnost na test skupu podataka iznosi 76%, dok su vrednosti za Top-1 i Top-3 tačnost 76% i 93%, respektivno.

VI. ZAKLJUČAK

Na osnovu postignutih rezultata i analize na skupu podataka STL10, pokušano je da se predstave praktični značaj i kvalitet ovih metoda. Za navedeni skup podataka, kao predstavnika značajnog skupa koji služi za upoređivanje modela samonadgledanog učenja, rezultati su vrlo zadovoljavajući i postignuta je tačnost od 76% na test skupu ciljnog zadatka.

Kao generalni problem tokom obuke na pseudo-zadatku istakao se problem invarijantnosti ugla. Deo slika u skupu podataka je invarijantan na ugao, odnosno slika može imati oba ugla - primer aviona koji je orijentisan sa leve na desnu stranu, ima apsolutno prihvatljivu i semantički ispravnu orijentaciju i ako je orijentisan sa desna na levo.

Važno je napomenuti da su dati skupovi podataka imali kanonski ugao od 0° što nije uvek slučaj, čime pseudo-zadatak rotacije može biti ugrožen u opštem slučaju primene.

Cilj ovog rada bio je da pokaže da samonadgledano učenje čak i na specifično dizajniranim arhitekturama dubokih neuronskih mreža postiže izvanredne rezultate u odnosu na broj označenih podataka za problem

klasifikacije slika. Daljim unapređivanjem i razvojem modela možemo očekivati da će biti u stanju da nadmaše rezultate nadgledanog učenja i postignu još kvalitetnije i značajnije rezultate koji se mogu koristiti u velikom broju aplikacija. Logično objašnjenje takvog stava jeste da se ovi modeli mogu obučavati na gotovo neograničenim skupovima podataka koji ne zahtevaju oznake, što u metodologiji nadgledanog učenja nije slučaj i iziskuje značajne ljudske i vremenske resurse.

LITERATURA

[1] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). Ieee.

[2] Jing, L., & Tian, Y. (2020). Self-supervised visual feature learning with deep neural networks: A survey. IEEE transactions on pattern analysis and machine intelligence, 43(11), 4037-4058.

[3] S. Gidaris, P. Singh & N. Komodakis. "Unsupervised representation learning by predicting image rotations." arXiv preprint arXiv:1803.07728 (2018)

[4] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[5] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

[6] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[7] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[8] Agarap, A. F. (2018). Deep learning using rectified linear units (relu). arXiv preprint arXiv:1803.08375.

[9] Ioffe, S., & Szegedy, C. (2015, June). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International conference on machine learning (pp. 448-456). PMLR.

[10] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), 1929-1958.

[11] Chollet, F., & others. (2015). Keras. GitHub. Retrieved from https://github.com/fchollet/keras

[12] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... & Zheng, X. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467.

[13] Coates, A., Ng, A., & Lee, H. (2011, June). An analysis of single-layer networks in unsupervised feature learning. In Proceedings of the fourteenth international conference on artificial intelligence and statistics (pp. 215-223). JMLR Workshop and Conference Proceedings.

[14] Zagoruyko, S., & Komodakis, N. (2016). Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. arXiv preprint arXiv:1612.03928.

ABSTRACT

Large amounts of labeled data are required to train deep neural networks to achieve good performance in the case of learning visual characteristics from images or videos in computer vision applications. To avoid the cost of collecting and labeling large datasets, a subset of unsupervised learning methods called self-supervised learning methods can be deployed. They manage to learn general visual characteristics of images and videos from unlabeled datasets. The paper implements a convolutional neural network that has the pseudo-task of recognizing which geometric transformation was applied to the image from the input, specifically - rotation. After training, using transfer learning techniques network was trained on a small subset of labeled data, for the task of image classification. On the STL10 dataset, an accuracy of 76% is achieved on the image classification downstream task..

SELF-SUPERVISED LEARNING USING A ROTATION PSEUDO-TASK ON A SPECIALLY DESIGNED DEEP NEURAL NETWORK ARCHITECTURE

Milutin Cerović