

IDŐBELI REKONSTRUKCIÓS
MÓDSZEREK AZ ÖNFELÜGYELT GÉPI
TANULÁSBAN

A DOKTORI ÉRTEKEZÉS TÉZISEI

MILACSKI ZOLTÁN ÁDÁM

Témavezető: Dr. habil. Lőrincz András
Információs Rendszerek Doktori Program
Vezető: Dr. Benczúr András
Informatika Doktori Iskola
Vezető: Dr. Csuhaj-Varjú Erzsébet
Mesterséges Intelligencia Tanszék
Informatikai Kar
ELTE Eötvös Loránd Tudományegyetem

2021 Március

A Mesterséges Intelligencia a Számítástudomány egyik tárgyköre, amely intelligens, emberekhez hasonlóan gondolkodó és cselekvő gépek kutatásával foglalkozik. Ezen belül a Gépi Tanulás [27, 30] ezt szemléltető példákban való tanulással éri el, megbecsülve az ismeretlen, mögöttes algoritmust, ami többféle módon is lehetséges, attól függően, hogy a példák mennyi információt tartalmaznak. A legegyszerűbb eset a Felügyelt Tanulás [39], ami emberek által összegyűjtött bemenet-célváltozó *tanító* párokkal dolgozik, és előbbiből az utóbbit próbálja megbecsülni. Bár az elmúlt években több lenyűgöző Felügyelt Tanulással kapcsolatos eredmény született, ezek gyakran rengeteg tanító párt igényelnek a megfelelő működéshez, amit monoton, költséges munka előállítani, és ily módon lehetőségeit behatárolja a humán erőforrás. Ezért előtérbe kerültek olyan alternatívák, mint a Felügyeletlen Tanulás [15, 16, 34, 40], ami pusztán a bemenetek ismeretét feltételezi: ezekre könnyebb szert tenni, hiszen nem igényelnek kézi párosítást. Az első ránézésre céltalannak tűnő feladat számos lehetőséggel kecsegtet. Speciálisan az Önfelügyelt Tanulás [37, 38] a bemenetekből automatikus módon generál ál-bemeneteket és ál-célváltozókat egy előre lefixált *mérési* függvény segítségével, majd ezekkel dolgozik a Felügyelt Tanulásbelihez hasonló módon, kvázi végtelen sok párral. Az Önfelügyelt Tanulás kitüntetett esetét képezik a Rekonstrukciós Feladatok: ezek direkt módon a bemenetet használják ál-célváltozóként, azaz a feladat a bemenet visszanyerése, ennek helyes megoldásához pedig magas és alacsony szintű változókat szükséges kinyerni az adatokból. A Rekonstrukciós Feladatok tipikusan két kategóriába sorolhatóak: Visszaállítási Feladatoknál [14, 31] az ál-bemenet valamilyen mérés; míg a Tömörítési Feladatoknál [17, 35, 43] mindkét ál-változó megegyezik a bemenettel, és veszteséges tömörítés történik. Szerencsére a kevés példával rendelkező Felügyelt Tanulási feladatok jól kondicionálttá tehetőek az Önfelügyelt Tanulással való közös megoldások kikényszerítésével. Ez az úgynevezett Félig Felügyelt Tanulás [25, 57], ami külön vagy együttes tanítással is véghezvihető (ezeket Átvitel Tanulásnak [44, 54] valamint Többfeladatú Tanulásnak [22, 46] szokás nevezni).

A probléma típusától függetlenül a Gépi Tanulási feladatok megoldása egy optimális függvény (leképezés) megtalálását célozza egy adott függvényhalmazon – az úgynevezett *hipotézistéren* [47] – belül, mely egy szintén adott célfüggvényt minimalizál. A Felügyelt és Önfelügyelt esetekben a célfüggvény a közelítés hibája: a tényleges és a becsült (ál-)célváltozó közötti eltérés átlagos nagysága az adott tanító példákra nézve. Azonban ez önmagában kevés, mert *túltanulás* [23, 29] következhet be: az optimális függvény kihasználhat olyan tulajdonságokat is, melyek csak az adott véges tanító halmazra igazak, de általában nem. Azaz a Gépi Tanulás célja az általánosítás is: a hibának kicsinek kell lennie korábban nem látott *teszt* példákra is. Ezt figyelembe véve, 2 fő hipotézistér típus terjedt el, melyek más-más esetekben, különféle szempontok szerint (értelmezhetőség, taníthatóság és sebesség) előnyösek. Egyrészt a paraméter nélküli optimalizálási függvények [26] hipotézisterei egyetlen, ember által is értelmezhető költségfüggvény minimalizálásából állnak, ami így nem tanítható és szükségszerűen előre megtervezett, továbbá lassú, iteratív megoldók használatát igényli még *teszt* fázisban is. Természetesen ezek alultanulást [36] visznek véghez, hiszen az adatoktól függetlenül jönnek létre, azonban jól általánosítanak. Másrészt a Mesterséges Mély Neurális Hálózatok [33, 41, 49] egyszerű, differenciálható, paraméteres leképezések összetett függvényei, amelyek taníthatóak és könnyedén kiszámolhatóak, azonban fekete dobozokként működnek. Ezek hipotézisterei gyakran egy konkrét függvény összes lehetséges paraméter-beállításait tartalmazzák, ez elég nagy halmaz a túltanulás előidézéséhez, ami ellen így további trükkök szükségesek. Újabban a két véglet ötvözetei is felbukkantak: egy előtanított (Mély Prioros Optimalizálás [24, 56]) vagy együttesen tanított (Iterációkra Bontott Mély Prioros Optimalizálás [28]) Mesterséges Mély Neurális Hálózat egy 2-tagú optimalizálási függvény változók szerint szeparálható megoldójában egy

ismeretlen tag *proximális operátoraként* [45] használható. Ezek az eljárások mindkét végtől merítenek jó tulajdonságokat: értelmezhetőek, taníthatóak és utóbbi kellően gyors is a valós idejű alkalmazásokhoz.

A Rekonstrukciós Feladatok széles köre és sok reménykeltő eredménye ellenére a szakirodalomban akadnak hiányosságok, például a szekvenciális adatok feldolgozásának tekintetében. Számos probléma és megoldás strukturálatlan vagy képi változókra korlátozódik, és nem taglalja ezek időbeli általánosításait, az idősorokat és a videókat. Továbbá a megoldások gyakran egyetlen feladatra specializálódnak. Ezek kedvezőtlen ismérvek, mivel az idő szerinti szekvenciális rendezés valamint a feladatok között fellépő korrelációk kiaknázása jobb változók tanulásának lehetőségével kecsegtet.

Disszertációmban 5 tézispontot mutatok be az iménti problémák orvoslására: időben terjesztem ki ismert Rekonstrukciós Feladatok definícióit és megoldásait, esetenként több feladatra együttesen. Kontribúcióm a fent kifejtett hipotézistér-családok törzsfejlődését követik.

Elsőként két új, paraméter nélküli optimalizálási függvényt mutatok be idősor adatok Zajtalanítási Feladataira, melyeknél a mérési függvény pusztán zajt ad hozzá a bemenethez. Módszereim képesek arra, hogy bármiféle tanulás nélkül pontosan detektáljanak olyan időpontokat, amelyek kiugróak vagy trendet váltanak. Eljárásaim csoportritka, időbeli prior regularizációs költségfüggvény tagokat használnak, ezáltal bizonyos minimumhelyeket tüntetnek ki: az anomáliák és a váltási pontok definíció szerint időben ritkák az idősorokon belül, ezt a tulajdonságot építem be előzetes tudásként az optimalizálási függvényekbe. A technikákat 2 Konvex Optimalizálási [20] függvénnyel (konkrétan, Regularizált Legkisebb Négyzeteken [32] keresztül) fogalmazom meg, így több elméleti garanciával is rendelkeznek (pl. egyértelmű globális minimumérték, erős dualitás és léteznek rájuk hatékony megoldók).

1. Tézis Egy Regularizált Legkisebb Négyzetek optimalizálási függvényt vezettem be idősor adatokra két új, időbeli, csoportritka prior költségfüggvénnyel: egy szakaszonként polinomiális és egy modell-alapú kiugrás detektáló priorral [2]. Az eljárás a Csoport LASSO [55], a Fuzionált LASSO [52] és a Robusztus Főkomponens Analízis [21] metódusokat kombinálja Konvex Optimalizáláson keresztül. Empirikus módon mutattam meg, hogy a módszer magasabb $F1$ értékeket ér el két rivális algoritmus-szal szemben egy robotkar kísérlet keretében, kiugró időpontok detektálásában.

2. Tézis Egy másik Regularizált Legkisebb Négyzetek optimalizálási függvényt készítettem idősor adatokra egy új súlyozott rekonstrukciós költségfüggvény definiálásával, ahol a súlyok természetes módon valószínűségeknek felelnek meg [1]. Az eljárás a Csoport LASSO, a Fuzionált LASSO és a Fuzionált Csoport LASSO [18] módszereken alapszik, Konvex Optimalizálás kihasználásával. Empirikusan demonstráltam, hogy a technika magasabb $F1$ értékeket ad az eredeti súlyozatlan Fuzionált LASSO és Fuzionált Csoport LASSO eljárásokhoz képest mesterségesen generált adatokon és valódi humán testpóz sorozatokon, váltási időpontok detektálásában.

Másodszor egy Mesterséges Mély Neurális Hálózatot vezetek be Tömörítési és Lineáris Visszaállítási (lineáris mérési függvényt használó) Feladatok kombinációjára, mely alkalmas videó osztályozásra való Átvitel Tanulást megelőző előtanításra. Eljárásom – hasonlóan a korábbi 2 tézispont-hoz – csoportritka feltevést használ, viszont ezúttal a ritkított teret az adatokból tanulom meg. Módszerem így gyorsabb és empirikusan jobb eredményekre vezet, cserébe a nemkonvex esetre való áttérés miatt elvesznek az elméleti garanciák és az interpretálhatóság.

3. Tézis Egy Mesterséges Mély Neurális Hálózatot (egy Konvolúciós Rekurrens Autoenkóder) terveztem videó adatokra két újonnan megadott (térbeli és élettartami) csoport k -ritka aktivációs függvény segítségével [4]. Ezek az ismert csoportstruktúra nélküli 1-ritka esetek változatai [42, 48], ahol a kiterjesztés a Csoport LASSO-hoz hasonlóan történik. Empirikusan validáltam két mesterségesen generált videó adatbázison, hogy osztályozásban, Átvitel Tanulás során, az eljárás jobb pontosságot ér el több ismert előtanító algoritmusnál, mint pl. a Zajtalanító Autoenkóderknél [51, 53], a csoportstruktúra nélküli 1-ritka megfelelőknél [42, 48] és a széles körben használt, ImageNet adatbázison Felügyelt módon betanított VGG-19 modellnél [50]. A minőségbeli javulások kevesebb címkézett tanítópélda esetén jelentősebbek voltak.

Harmadszor két, Iterációkra Bontott Mély Prioros Optimalizálási függvényt javaslok, melyek több, képeken és videókon definiált Lineáris Visszaállítási Feladatot képesek egyszerre megoldani. Az első 3 tézisponttal ellentétben ezek lehetővé teszik, hogy a csoportritkaság által megengedetteknél általánosabb prior feltevéseket tanuljak az adatokból. Eljárásaim ötvözik az optimalizálási függvények és a Mesterséges Mély Neurális Hálózatok előnyeit.

4. Tézis Egy Iterációkra Bontott Mély Prioros Regularizált Legkisebb Négyzetek optimalizálási függvényt ismertettem Multiplikátorok Váltakozó Irányú Módszere [19] megoldóval képek Lineáris Visszaállítására, ahol a prior költségfüggvény implicit módon került definiálásra egy Mesterséges Mély Neurális Hálózat (egy Konvolúciós Autoenkóder) segítségével [3]. A procedúra az Egy Háló Mindent Megold [24] eljárás új, együttesen tanítható változataként és az Iterációkra Bontott Mély Prioros Optimalizálás [28] többfeladatú általánosításaként is értelmezhető. Empirikusan bizonyítottam az MS-Celeb-1M és ImageNet adatbázisokon, hogy a javasolt módszer szignifikánsan magasabb Csúcs Jel-Zaj Viszony értékeket ér el, mint az eredeti Egy Háló és a népszerű 2-D Wavelet ritkaság algoritmusok, méghozzá jóval kevesebb iteráció alatt.

5. Tézis A 4. Tézispontbeli optimalizálási függvényemet videó adatokra terjesztettem ki 2 kontribúció hozzáadásával: a Konvolúciós Autoenkóder priort lecseréltem egy Kétirányú Konvolúciós Rekurrens Neurális Hálózatra, valamint a megoldó eddig fixált adatfeldolgozó lépését is taníthatóvá tettem egy második ugyanilyen hálózat hozzáadásával. [5]. A teljes architektúrát Két Háló Mindent Megoldnak kereszteltem. Empirikusan igazoltam 2 mesterségesen generált és a valódi UCF-101 videó adatbázison, hogy a javasolt módosításaim Lineáris Visszaállítás során egyenként javítják a Csúcs Jel-Zal Viszony értékeket a 4. tézispontbeli Egy Háló változathoz és a közismert 3-D Wavelet ritkasághoz képest; valamint, hogy a 2 változtatás együttesen még jobb eredményekre vezet.

Kísérleti eredményeim teljes mértékben reprodukálhatóak a leírások és a mellékelt forráskódok alapján, melyek további vizsgálatok alapjaként szolgálhatnak.

A fenti tézispontok mellett más, lazán kapcsolódó publikációim is születtek. Ezek elő- illetve kiegészítő tanulmányoknak tekintendők, amik csoportmunka keretében készültek, így az önálló tézispontok szintjéhez kevésnek bizonyultak. Társszerzőimmel egyéb Rekonstrukciós Feladatokat vizsgáltunk idősor és videó adatokon, melyeket ritka optimalizálási függvényekkel [12, 13] és Mesterséges Mély Neurális Hálózatokkal (szintén ritkaság [10, 11]

vagy más regularizációs technikák [8, 9] segítségével) oldottunk meg. Ezenfelül részt vettem egy szubmoduláris függvénycsaládot Mesterséges Mély Neurális Hálózatokkal összekapcsoló matematikai bizonyítás kidolgozásában [6]. Végül statisztikailag elemeztem egy Felügyelt Tanuláshoz humán annotátorokat kiképző tanfolyam jótékony hatását [7].

A TÉZISEK ALAPJÁUL SZOLGÁLÓ PUBLIKÁCIÓK

- [1] András Lőrincz, Máté Csákvári, Áron Fóthi, Zoltán Ádám Milacski, András Sárkány, and Zoltán Tóser. “Towards Reasoning Based Representations: Deep Consistence Seeking Machine”. In: *Cognitive Systems Research* 47 (2018), pp. 92–108.
- [2] Zoltán Ádám Milacski, Marvin Ludersdorfer, András Lőrincz, and Patrick Van Der Smagt. “Robust Detection of Anomalies via Sparse Methods”. In: *International Conference on Neural Information Processing (ICONIP)*. Springer. 2015, pp. 419–426.
- [3] Zoltán Ádám Milacski, Barnabás Póczos, and András Lőrincz. “Differentiable Unrolled Alternating Direction Method of Multipliers for OneNet”. In: *British Machine Vision Conference (BMVC)*. BMVA. 2019.
- [4] Zoltán Ádám Milacski, Barnabás Póczos, and András Lőrincz. “Group k -Sparse Temporal Convolutional Neural Networks: Unsupervised Pre-training for Video Classification”. In: *IEEE International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2019.
- [5] Zoltán Ádám Milacski, Barnabás Póczos, and András Lőrincz. “Video-OneNet: Bidirectional Convolutional Recurrent OneNet with Trainable Data Steps for Video Processing”. In: *International Conference on Machine Learning (ICML)*. 2020.

A DOLGOZAT TÉMÁJÁBAN MEGJELENT TOVÁBBI PUBLIKÁCIÓK

- [6] Kristóf Bérczi, Erika Renáta Bérczi-Kovács, András Lőrincz, and Zoltán Ádám Milacski. “Facility location functions are deep submodular functions”. In: *EGRES Quick-Proofs* 01 (2019).
- [7] Kinga Bettina Faragó, Zoltán Ádám Milacski, András Németh, and András Lőrincz. “Enhancing Crowdsourced Applications via Incorporated Practice Sessions”. In: *International Workshop on Gamification for Information Retrieval, European Conference on Information Retrieval (GamifIR@ ECIR)*. 2015, pp. 41–43.
- [8] Changhee Han, Leonardo Rundo, Kohei Murao, Zoltán Ádám Milacski, Kazuki Umemoto, Evis Sala, et al. “GAN-based Multiple Adjacent Brain MRI Slice Reconstruction for Unsupervised Alzheimer’s Disease Diagnosis”. In: *arXiv:1906.06114* (2019).
- [9] Changhee Han, Leonardo Rundo, Kohei Murao, Tomoyuki Noguchi, Yuki Shimahara, Zoltán Ádám Milacski, et al. “MADGAN: unsupervised Medical Anomaly Detection GAN using multiple adjacent brain MRI slice reconstruction”. In: *arXiv:2007.13559* (2020).
- [10] András Lőrincz, Zoltán Ádám Milacski, Balázs Pintér, and Anita Lilla Verő. “Columnar Machine: Fast estimation of structured sparse codes”. In: *Biologically Inspired Cognitive Architectures* 15 (2016), pp. 19–33.
- [11] András Lőrincz, András Sárkány, Zoltán Ádám Milacski, and Zoltán Tóser. “Estimating Cartesian Compression via Deep Learning”. In: *International Conference on Artificial General Intelligence (AGI)*. Springer. 2016, pp. 294–304.

- [12] Zoltán Ádám Milacski, Kinga Bettina Faragó, Áron Fóthi, Viktor Varga, and András Lőrincz. “Declarative Description: The Meeting Point of Artificial Intelligence, Deep Neural Networks and Human Intelligence”. In: *Workshop on Explainable Artificial Intelligence, International Joint Conference on Artificial Intelligence (IJCAI XAI)*. 2018, pp. 97–103.
- [13] Tamás Tettamanti, Zoltán Ádám Milacski, András Lőrincz, and István Varga. “Iterative Calibration Method for Microscopic Road Traffic Simulators”. In: *Periodica Polytechnica Transportation Engineering* 43.2 (2015), pp. 87–91.

IRODALOMJEGYZÉK

- [14] Simon Arridge, Peter Maass, Ozan Öktem, and Carola-Bibiane Schönlieb. “Solving inverse problems using data-driven models”. In: *Acta Numerica* 28 (2019), pp. 1–174.
- [15] Horace B Barlow. “Unsupervised learning”. In: *Neural Comput.* 1.3 (1989), pp. 295–311.
- [16] Yoshua Bengio, Aaron C Courville, and Pascal Vincent. “Unsupervised Feature Learning and Deep Learning: A Review and New Perspectives”. In: *CoRR, abs/1206.5538* 1 (2012), p. 2012.
- [17] Yoshua Bengio, Aaron Courville, and Pascal Vincent. “Representation Learning: A Review and New Perspectives”. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 35.8 (2013), pp. 1798–1828.
- [18] Kevin Bleakley and Jean-Philippe Vert. “The Group Fused LASSO for Multiple Change-Point Detection”. In: *arXiv:1106.4199* (2011).
- [19] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. “Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers”. In: *Found. Trends® Mach. Learn.* 3.1 (2011), pp. 1–122.
- [20] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge Univ. Press, 2004.
- [21] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. “Robust Principal Component Analysis?” In: *J. ACM* 58.3 (2011), pp. 1–37.
- [22] Rich Caruana. “Multitask Learning”. In: *Mach. Learn.* 28.1 (1997), pp. 41–75.
- [23] Rich Caruana, Steve Lawrence, and C Lee Giles. “Overfitting in Neural Nets: Backpropagation, Conjugate Gradient, and Early Stopping”. In: *Adv. NeurIPS*. 2001, pp. 402–408.
- [24] JH. R. Chang, Chun-Liang Li, Barnabás Póczos, BVK Vijaya Kumar, and Aswin C Sankaranarayanan. “One Network to Solve Them All—Solving Linear Inverse Problems Using Deep Projection Models”. In: *IEEE ICCV*. 2017, pp. 5888–5897.
- [25] Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien. “Semi-Supervised Learning”. In: *IEEE Trans. Neural Netw.* 20.3 (2009), pp. 542–542.
- [26] George B Dantzig. “The Nature of Mathematical Programming”. In: *Mathematical Programming Glossary* (2010).
- [27] Ayon Dey. “Machine Learning Algorithms: A Review”. In: *Int. J. Comput. Sci. Inf. Technol.* 7.3 (2016), pp. 1174–1179.
- [28] Steven Diamond, Vincent Sitzmann, Felix Heide, and Gordon Wetzstein. “Unrolled Optimization with Deep Priors”. In: *arXiv:1705.08041* (2017).
- [29] Tom Dietterich. “Overfitting and Undercomputing in Machine Learning”. In: *ACM Comput. Surv.* 27.3 (1995), pp. 326–327.

- [30] Pedro Domingos. "A Few Useful Things to Know About Machine Learning". In: *Commun. ACM* 55.10 (2012), pp. 78–87.
- [31] Heinz Werner Engl, Martin Hanke, and Andreas Neubauer. *Regularization of Inverse Problems*. Vol. 375. Springer Science & Business Media, 1996.
- [32] Gene H Golub, Per Christian Hansen, and Dianne P O’Leary. "Tikhonov Regularization and Total Least Squares". In: *SIAM J. Matrix Anal. Appl.* 21.1 (1999), pp. 185–194.
- [33] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press, 2016.
- [34] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. "Unsupervised Learning". In: *The Elements of Statistical Learning*. Springer, 2009, pp. 485–585.
- [35] Geoffrey E Hinton and Ruslan R Salakhutdinov. "Reducing the Dimensionality of Data with Neural Networks". In: *Science* 313.5786 (2006), pp. 504–507.
- [36] H Jabbar and Rafiqul Zaman Khan. "Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study)". In: *Comput. Sci. Commun. Instrum. Devices* (2015).
- [37] Longlong Jing and Yingli Tian. "Self-supervised Visual Feature Learning with Deep Neural Networks: A Survey". In: *IEEE Trans. Pattern Anal. Mach. Intell.* (2020).
- [38] Alexander Kolesnikov, Xiaohua Zhai, and Lucas Bayer. "Revisiting Self-Supervised Visual Representation Learning". In: *IEEE CVPR*. 2019, pp. 1920–1929.
- [39] Sotiris B Kotsiantis, I Zaharakis, and P Pintelas. "Supervised Machine Learning: A Review of Classification Techniques". In: *Emerg. Artif. Intell. Appl. Comput. Eng.* 160 (2007), pp. 3–24.
- [40] Martin Långkvist, Lars Karlsson, and Amy Loutfi. "A review of unsupervised feature learning and deep learning for time-series modeling". In: *Pattern Recognit. Lett.* 42 (2014), pp. 11–24.
- [41] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". In: *Nature* 521.7553 (2015), p. 436.
- [42] Alireza Makhzani and Brendan J Frey. "Winner-Take-All Autoencoders". In: *Adv. NeurIPS*. 2015, pp. 2791–2799.
- [43] Andrew Ng et al. "Sparse autoencoder". In: *CS294A Lecture Notes* 72.2011 (2011), pp. 1–19.
- [44] Sinno Jialin Pan and Qiang Yang. "A Survey on Transfer Learning". In: *IEEE Trans. Knowl. Data Eng.* 22.10 (2009), pp. 1345–1359.
- [45] Neal Parikh and Stephen Boyd. "Proximal Algorithms". In: *Found. Trends Optim.* 1.3 (2014), pp. 127–239.
- [46] Sebastian Ruder. "An Overview of Multi-Task Learning in Deep Neural Networks". In: *arXiv:1706.05098* (2017).
- [47] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 2002.
- [48] Eder Santana, Matthew S Emigh, Pablo Zegers, and Jose C Principe. "Exploiting Spatio-Temporal Structure with Recurrent Winner-Take-All Networks". In: *IEEE Trans. Neural Netw. Learn. Syst.* 29.8 (2018), pp. 3738–3746.
- [49] Jürgen Schmidhuber. "Deep Learning in Neural Networks: An Overview". In: *Neural Netw.* 61 (2015), pp. 85–117.
- [50] Karen Simonyan and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition". In: *arXiv:1409.1556* (2014).

- [51] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *J. Mach. Learn. Res.* 15.1 (2014), pp. 1929–1958.
- [52] Robert Tibshirani, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. "Sparsity and smoothness via the Fused LASSO". In: *J. Royal Stat. Soc. B* 67.1 (2005), pp. 91–108.
- [53] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. "Extracting and Composing Robust Features with Denoising Autoencoders". In: *ICML*. ACM. 2008, pp. 1096–1103.
- [54] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. "A survey of transfer learning". In: *J. Big Data* 3.1 (2016), p. 9.
- [55] Ming Yuan and Yi Lin. "Model Selection and Estimation in Regression with Grouped Variables". In: *J. Royal Stat. Soc. B* 68.1 (2006), pp. 49–67.
- [56] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. "Learning Deep CNN Denoiser Prior for Image Restoration". In: *IEEE CVPR*. 2017, pp. 3929–3938.
- [57] Xiaojin Zhu and Andrew Goldberg. *Introduction to Semi-Supervised Learning*. Morgan & Claypool Publ., 2009.