UNIVERSITATEA NAȚIONALĂ DE ȘTIINȚĂ ȘI TEHNOLOGIE POLITEHNICA BUCUREŞTI

Şcoala Doctorală de Electronică, Telecomunicaţii şi Tehnologia Informaţiei

Decizie nr. 119 **din** 26-10-2023

REZUMAT TEZĂ DE DOCTORAT

Ana-Antonia NEACȘU

METODE ROBUSTE DE ÎNVĂȚARE PROFUNDĂ INSPIRATE DIN ALGORITMI DE PROCESARE DE SEMNAL

ROBUST DEEP LEARNING METHODS INSPIRED BY SIGNAL PROCESSING ALGORITHMS

COMISIA DE DOCTORAT

Prof. Dr. Ing. Mihai CIUC Univ. Națională de Științe și Tehnologie Politehnica București	Președinte
Prof. Dr. Ing. Corneliu BURILEANU Univ. Națională de Științe și Tehnologie Politehnica București	Co-îndrumător
Prof. Dr. Ing. Jean-Christophe Pesquet CentraleSupélec, Université Paris-Saclay	Co-îndrumător
Prof. Dr. Ing. Nicu SEBE University of Trento, Italy	Recenzor
Prof. Dr. Ing. Corneliu RUSU Univ. Tehnica din Cluj-Napoca	Recenzor
Prof. Dr. Ing. Daniela TĂRNICERIU Univ. Tehnică "Gh. Asachi" Iași	Recenzor
Dr. Ing. Jean-Philippe OVARLEZ Research Director at ONERA, Université Paris-Saclay, France	Examinator
Dr. Ing. Frank MAMALET Senior Expert in Artificial Intelligence at IRT Saint Exupéry, Toulouse, France	Examinator

BUCUREŞTI 2023

Cuprins

Capitolul 1

Introducere

1.1 Context

Recent, metodele de învățare automată au devenit instrumente omniprezente într-o gamă largă de sarcini, datorită capacității lor de a rezolva o mare varietate de probleme, de la regresii simple la clasificări multimodale complexe. Aceste metode stau în centrul *Inteligenței Artificiale* (IA). IA reprezintă minunea tehnologiei contemporane și este folosită cu succes într-un număr tot mai mare de domenii care afectează viețile noastre, cum ar fi medicina [20], conducerea autonomă [27], procesarea limbajului natural [42], interacțiunea om-calculator (HCI) [33], etc. Cu toate acestea, rețelele neuronale adânci, probabil cele mai puternice metode, ridică provocări în ceea ce privește problematica implementării în timpul fazei de învățare. Mai mult, ele par a fi cutii negre al căror caracter robust nu este întotdeauna bine controlat [14, 31].

Dezvoltarea unui sistem bazat pe IA de încredere este esențială pentru a asigura că sistemele inteligente pot fi folosite cu încredere în luarea deciziilor critice fără a compromite standardele etice.

Pentru a atinge acest obiectiv, o problemă critică de abordat în dezvoltarea aplicațiilor în viața reală folosind rețele neurale este evaluarea și controlul corect al robusteții acestora împotriva posibilelor atacuri adversare.

Intrările adversare reprezintă date de intrare malițioase care pot păcăli modelele de învățare automată. Conceptul a fost evidențiat în [38], unde autorii au arătat că modificarea ușoară a datelor de intrare care au fost corect clasificate de rețea poate duce la o clasificare greșită [23, 3, 40, 18].

Trebuie subliniat faptul că intrările adverse nu sunt neapărat create artificial cu intenția de a sabota sistemul. Ele pot apărea și în mod natural sub diferite forme și pot afecta grav performanța aplicațiilor din viața reală bazate pe modele pre-antrenate [29]. O analiză mai bună a proprietăților de stabilitate ale rețelelor neurale poate fi văzută ca primul pas către o mai bună înțelegere a principiilor matematice care guvernează funcționalitățile lor.

Scopul principal al acestei teze este să proiecteze noi metode pentru antrenarea rețelelor neurale sigure, dar cu performanță ridicată. Rezultatele matematice recente arată că devine mai ușor să controlezi stabilitatea rețelelor neurale introducând con-

strângeri potrivite asupra ponderilor lor. Cu toate acestea, aceasta necesită gestionarea unor constrângeri care nu sunt neapărat convexe în faza de antrenare a rețelei neurale. În acest sens, am *proiectat constrângeri* pe care le-am folosit ulterior în procesul de antrenare pentru a asigura robustețea rețelei neurale. După cum este evidențiat în [16], *comportamentul Lipschitz* al rețelei este strâns corelat cu robustețea sa împotriva atacurilor adverse. Această constantă ne permite să limităm superior perturbarea de ieșire cunoscând magnitudinea celei de intrare, pentru o anumită metrică [36]. Controlul acestei constante duce la o soluție fezabilă pentru evaluarea efectului atacurilor adverse dacă este calculată cu precizie. Cu toate acestea, calcularea exactă a constantei Lipschitz chiar și pentru o rețea neurală superficială este o problemă non-polinomială *NP-hard*. Prin urmare, dificultatea principală constă în a*găsi modalități de a o aproxima cât mai strâns posibil*. Recent, au fost propuse mai multe metode pentru antrenarea rețelelor Lipschitziene, care se încadrează în două categorii principale. Abordările de regularizare includ dubla retropropagare [12] sau aplicarea penalizării asupra Jacobianei rețelei [17], care impune constrângeri Lipschitz locale, dar nu impun constrângerea globală asupra rețelei. O altă abordare constă în impunerea unor constrângeri asupra arhitecturii rețelei, astfel încât să limiteze norma spectrală a fiecărui strat [40] [10]. Cu toate că au o complexitate computațională crescută, aceste metode asigură o rețea Lipschitziană. În [11], au fost propuse rezultate noi care duc la aproximări precise ale constantei Lipschitz pentru rețelele neurale pozitive. Aceste rezultate preliminare au servit drept punct de plecare pentru propunerea de metode eficiente pentru proiectarea de rețele neurale sigure.

După stabilirea întregii baze matematice, ne concentrăm apoi asupra *construirii de noi arhitecturi de rețele neurale* bazate pe filozofia menționată anterior. O parte importantă a lucrării prezentate în această teză constă în *dezvoltarea de metode eficiente de optimizare pentru învățarea supervizată a rețelelor neurale*. Analizăm opțiunile posibile pentru structura rețelei, având în vedere diferitele clase de algoritmi de optimizare iterativă existenți. Pentru a gestiona constrângerile de stabilitate, acordăm o atenție deosebită *metodelor proximale*, care oferă instrumente puternice pentru optimizare într-un context la scară mare. Studiem cum asigurarea *robustezii afectează performanța globală* a sistemelor de învățare și încercăm să obținem un *echilibru bun între robustețe și acuratețe*.

Un aspect foarte important în toate cercetările exploratorii este *validarea rezultatelor teoretice într-un context real de aplicație*. Unele dintre modelele antrenate cu garanții de stabilitate sunt testate în contexte din viața reală pentru a arăta versatilitatea soluțiilor proiectate. Măsurăm apoi influența asupra performanței sistemului și *comparăm* rezultatele obținute cu cele generate cu arhitecturi clasice, precum și cu alte strategii de apărare.

1.2 Impact și aplicabilitate

Această teză contribuie la domeniul învățării automate încercând să ofere un răspuns la întrebarea fundamentală:

Obiectivul este de a furniza garanții de robustețe demonstrate matematic, de a dezvolta software-ul asociat și de a-l face disponibil public. Un alt aspect important al acestei teze este focusul pe aplicații bazate pe semnale audio și fiziologice care au utilizare directă în dezvoltarea de tehnologii inovatoare și pot beneficia direct o varietate de produse destinate consumatorilor.

Într-o manieră mai generală, abordând conceptul de Rețele Neurale Sigure, această teză contribuie la starea cunoștințelor în domeniul inteligenței artificiale, valorificând cele mai recente rezultate de cercetare în domeniul optimizării. Dezvoltarea de noi metode care pot fi folosite pentru a face sistemele de învățare mai robuste și explicabile va deschide noi perspective în ceea ce privește progresul tehnologic sigur și controlat.

1.3 Principalele contribuții

Primele contribuții ale tezei apar în Capitolul 3:

(i) Propunem un sistem robust de recunoaștere automată a gesturilor în timp real bazat pe semnale sEMG. Robustețea este asigurată utilizând un algoritm de învățare nou pentru antrenarea rețelelor neurale.

(ii) Arătăm că se poate atinge un echilibru bun între acuratețe și robustețe. Pentru a face acest lucru, antrenăm sistemul sub constrângeri de normă spectrală, permițându-ne să controlăm fin constanta sa Lipschitz. O constantă Lipschitz strânsă este estimată eficient concentrându-ne pe rețele neurale cu greutăți strict pozitive, așa cum se face și în [8].

(iii) Demonstrăm performanța arhitecturii finale în experimente din viața reală, unde arătăm că modelul robust propus depășește pe cele antrenate în mod convențional.

(iv) Analizăm modul în care sistemul nostru se comportă atunci când intrarea este afectată de diferite niveluri de zgomot, simulând perturbații care pot apărea în scenarii reale.

(v) Arătăm validitatea soluției noastre prin experimente pe mai multe seturi de date publice de gesturi sEMG.

Capitolul 4 include următoarele contribuții principale.

(i) Inspirați de filtrele MIMO, introducem o nouă clasă de rețele neurale, care pot fi văzute ca o soluție intermediară între CNN-uri și FCN-uri.

(ii) Propunem o strategie de antrenare constrânsă, care ne permite să controlăm constanta Lipschitz a rețelei pentru a asigura robustețea sa la zgomot adversar.

(iii) Prezentăm o nouă arhitectură (RCFF-Net), care operează în domeniul valorilor complexe, pentru care derivăm limite strânse ale constantei Lipschitz.

(iv) Dezvoltăm o strategie de învățare constrânsă pentru a antrena structura propusă în timp ce controlăm constanta Lipschitz globală.

(v) Ambele arhitecturi ACNN și RCFF sunt evaluate în sarcini de eliminare a zgomotului din semnalele audio, demonstrând că soluția noastră nu se limitează la problemele de clasificare.

Contribuțiile din Capitolul 5 sunt menționate mai jos.

(i) Introducem rețelele ABBA, o clasă nouă de rețele neurale (aproape) strict pozitive, care posedă o serie de proprietăți atractive.

(ii) Arătăm că putem pune orice rețea cu semne arbitrare într-o formă ABBA. Arătăm că această proprietate este valabilă atât pentru rețelele complet conectate, cât și pentru cele convoluționale.

(iii) Derivăm teoreme de aproximare universală pentru rețelele care prezintă straturi ponderate strict pozitiv.

(iv) Prezentăm o metodă pentru controlul eficient al constantei Lipschitz a rețelelor ABBA. Această strategie de control se aplică atât în cazul rețelelor complet conectate, cât și în cazul celor convoluționale.

(v) Experimentele numerice efectuate pe seturi de date standard de imagini evidențiază performanța excelentă a rețelelor ABBA pentru modele mici. În mod remarcabil, acestea prezintă îmbunătățiri semnificative atât în performanță, cât și în robustețe, în comparație cu rețelele cu ponderi exclusiv strict pozitive. Mai mult, demonstrăm că rețelele ABBA sunt competitive cu rețelele robuste care prezintă ponderi cu semne arbitrare, antrenate cu tehnici de moderne.

1.4 Publicații

Aricole de jurnal trimise

- A. Neacșu, J.-C. Pesquet, V. Vasilescu and C.Burileanu, "*ABBA Neural Networks: Coping with Positivity, Expressivity, and Robustness*", submitted to SIAM Journal on Mathematics of Data Science (SIMODS), 2023.

Aricole de jurnal publicate

- A. Neacșu, J.-C. Pesquet and C.Burileanu, "*EMG-Based Automatic Gesture Recognition Using Lipschitz-Regularized Neural Networks*", accepted for publication in ACM Transactions on Intelligent Systems and Technology (TIST), 2023.

- N Lassau, S. Ammari, E. Chouzenoux, A. Neacșu et al. *"Integrating deep learning CT-scan model, biological and clinical variables to predict severity of COVID-19 patients"*,in Nature Communication 12, 634 (2021), https://doi.org/10.1038/s41467-020-20657-4

Conferințe

- C. Andronache, M. Negru, I. Bădiţoiu, G. Cioroiu, A. Neacsu and C. Burileanu, "*Automatic Gesture Recognition Framework Based on Forearm EMG Activity*", in Proc. 45th International Conference on Telecommunications and Signal Processing (TSP), Prague, Czech Republic, 2022, pp. 284-288, doi: 10.1109/TSP55681.2022.9851314.
- A. Neacşu, R. Ciubotaru, J. -C. Pesquet and C. Burileanu, "*Design of Robust Complex-Valued Feed- Forward Neural Networks*", in Proc. 30th European Signal Processing Conference (EUSIPCO), Belgrade, Serbia, 2022, pp. 1596-1600, doi: 10.23919/EUSIPCO55093.2022.9909696.
- A. Neacșu, K. Gupta, J. -C. Pesquet and C. Burileanu, "*Signal Denoising Using a New Class of Robust Neural Networks*" in Proc. of 28th European Signal Processing Conference (EUSIPCO), Amsterdam, Netherlands, 2021, pp. 1492-1496, doi: 10.23919/Eusipco47968.2020.9287630.
- V. Vasilescu, A. Neacşu, E. Chouzenoux, J. -C. Pesquet and C. Burileanu, "*A Deep Learning Approach For Improved Segmentation Of Lesions Related To Covid-19 Chest CT Scans*", in Proc. IEEE 18th Int. Sym. on Biomedical Imaging (ISBI), Nice, France, 2021, pp. 635-639, doi: 10.1109/ISBI48211.2021.9434139.
- A. Neacșu, J.-C. Pesquet, and C. Burileanu, "*Accuracy-robustness trade-off for positively weighted neural networks*", in Proc. IEEE International Conference on Acoustics and Speech Signal Process. (pp. 8389–8393). Barcelona, Spain, 2020, doi: 10.1109/ICASSP40776.2020.9053803.
- C. Andronache, M. Negru, A. Neacsu, G. Cioroiu, A. Radoi and C. Burileanu, "*Towards extending real-time EMG-based gesture recognition system*", in Proc. 43rd International Conference on Telecommunications and Signal Processing (TSP), Milan, Italy, 2020, pp. 301-304,
  doi: 10.1109/TSP49548.2020.9163481.

1.5 Teză în co-tutelă

Colaborarea stă la baza progresului științific și a inovației. În lumea interconectată de astăzi, importanța eforturilor colaborative nu poate fi subestimată, în special în domeniul cercetării academice. Această teză este rezultatul unei colaborări în co-tutelă între Universitatea Politehnică din București și CentraleSupélec, Școala de Studii de Inginerie a Universității Paris Saclay. Această teză a oferit o oportunitate remarcabilă pentru a promova cooperarea și schimbul de cunoștințe între cele două instituții de înaltă calitate.

1.6 Structura tezei

Restul tezei este organizat după cum urmează. În Capitolul 2, prezentăm o privire de ansamblu asupra strategiilor existențe de atac și defensivă împotriva perturbațiilor

adversare existente. În Secțiunea 2.1 stabilim conceptul de robustețe în contextul rețelelor neurale, în timp ce în Secțiunea 2.2 introducem notația matematică utilizată pe parcursul capitolului. Prezentăm cele mai utilizate scenarii ale modelelor de amenințare (Secțiunea 2.3) și apoi descriem mecanismele de atac atât de tip cutie albă, cât și de tip cutie neagră în Secțiunea 2.4.Încheiem capitolul evidențiind diferitele strategii de apărare în Secțiunea 2.5.

În Capitolul 3, prezentăm un mecanism robust pentru antrenarea rețelelor neurale strict pozitive în contextul recunoașterii automate a gesturilor bazate pe semnale sEMG. În Secțiunea 3.1 punem bazele înțelegerii electromiografiei și subliniem relevanța sa în contextul recunoașterii gesturilor. În continuare, în Secțiunea 3.2 introducem abordări inovatoare pentru a îmbunătăți robustețea rețelelor neurale complet conectate. Secțiunea 3.3 detaliază apoi tehnicile de optimizare esențiale pentru metodele noastre propuse, variante ale cărora vor fi folosite în restul acestei lucrări. Trecând la implementarea practică, Secțiunea 3.4 oferă perspective asupra cadrului experimental considerat pentru sarcina noastră. Capitolul culminează cu Secțiunea 3.5, unde validăm pe larg robustețea modelelor noastre propuse. În final, încheiem acest capitol sumarizând principalele noastre concluzii și implicații în Secțiunea 3.6.

În Capitolul 4 ne îmbarcăm într-o călătorie pentru a îmbunătăți calitate semnalelor audio folosind rețele neurale robuste. Începând cu Secțiunea 4.1 introducem prima arhitectură inovatoare pe care o propunem în această teză. Apoi explorăm, în Secțiunea 4.1.1, un pas critic în a traversa decalajul dintre aceste paradigme ale rețelelor neurale: utilizarea straturilor complet conectate și convoluționale. Secțiunea 4.1.2 intră în strategiile de optimizare folosite pentru antrenarea modelelor noastre propuse, evidențiind nucleul metodologiei noastre. Aplicațiile practice dezvoltate în Secțiunea 4.1.3 oferă o examinare detaliată a performanței modelelor în scenarii de eliminare a zgomotului din semnale. A doua parte a capitolului, începând cu Secțiunea 4.2, introduce o nouă clasă de rețele (RCFF) care operează în domeniul complex. Fundamentele și ideile teoretice sunt prezentate în Secțiunile 4.2.1-4.2.2, unde elucidăm bazele matematice ale mecanismelor robuste de antrenare, apoi detaliem implementarea lor în Secțiunea 4.2.3. Ulterior, prezentăm rezultatele empirice ale aplicării RCFF-Net la probleme de eliminare a zgomotului din semnalele audio în Secțiunea 4.2.4. În cele din urmă, concluzionăm acest capitol sumarizând principalele noastre constatări și implicații în Secțiunea 4.3.

În Capitolul 5, introducem o clasă revoluționară de rețele neurale cunoscute sub numele de Rețele Neurale ABBA, proiectate să abordeze probleme legate de pozitivitate, expresivitate și robustețe. Începem cu Secțiunea 5.1, oferind o privire de ansamblu asupra provocărilor pe care noile noastre rețele ABBA își propun să le abordeze. Oferim context în Secțiunea 5.2, examinând peisajul existent al soluțiilor de rețele neurale și evidențiind contribuțiile unice ale rețelelor ABBA. Nucleul capitolului nostru se desfășoară în Secțiunea 5.3, unde descriem fundamentele arhitecturale și trăsăturile cheie ale acestei clase inovatoare de rețele neurale. Ulterior, în Secțiunea 5.4, extindem aplicabilitatea rețelelor ABBA la cazul convoluțional, evidențiind adaptabilitatea acestei abordări în diverse arhitecturi de rețele. O privire detaliată asupra metodelor și tehnicilor de antrenare care asigură stabilitatea Lipschitz este prezentată în Secțiunea 5.5. Secțiunea

5.6 servește ca nucleu empiric al acestui capitol, unde efectuăm evaluări cuprinzătoare pentru a valida performanța și eficacitatea rețelelor ABBA în diverse scenarii de clasificare. În Secțiunea 5.7, rezumăm principalele noastre constatări, idei și implicări ale cercetării noastre.

În final, în Capitolul **??**, tragem concluziile finale ale acestei teze, urmate de o scurtă descriere a unor perspective.

Capitolul 2

Privire de ansamblu asupra atacurilor și apărărilor adversative

Acest capitol prezintă o privire de ansamblu asupra progreselor actuale în domeniul robusteții rețelelor neurale împotriva perturbațiilor adversative. Definim conceptul de atacuri adversative și explorăm aspectele celor mai eficiente strategii de atac. Studierea atacurilor deliberate create în domeniul învățării automate este crucială, deoarece permite identificarea vulnerabilităților modelelor și îmbunătățirea robusteții acestora.

2.1 Robustețea rețelelor neurale

Această secțiune subliniază necesitatea înțelegerii și îmbunătățirii rezistenței rețelelor neurale la intrări adversative, explorând conceptul de robustețe, crearea perturbațiilor și strategiile pentru a reduce impactul acestora.

2.2 Definiții și notații

În această secțiune, sunt introduse principalele notații utilizate pe parcursul capitolului.

2.3 Modele de amenințare

Această secțiune discută opțiunile posibile pentru modelele de amenințare, în funcție de obiectivul și nivelul de acces la modelul original, acestea putând fi încadrate în mai multe categorii, după cum urmează. În funcție de obiectivul adversarului, atacurile pot fi *țintite* sau *netințite*. În plus, în funcție de nivelul de acces pe care atacatorul îl are asupra modelului victimă, se disting trei categorii distincte de atacuri: atacuri *cutie neagră* (*black-box*), atacuri *cutie albă* (*white-box*) și atacuri *cutie gri* (*gray-box*).

2.4 Mecanisme de atac

În această secțiune, detaliem principalii algoritmi folosiți pentru generarea de exemple adversative în toate cele trei contexte. Luăm în considerare în special metodele de evitare, deoarece acestea sunt mai frecvente.

2.5 Strategii de apărare

Deoarece există multe moduri în care un adversar poate exploata vulnerabilitățile modelului, au fost dezvoltate strategii de apărare pentru a atenua această problemă de robustețe. Această secțiune prezintă principalele direcții în acest domeniu.

2.6 Concluzie

Acest capitol prezintă o privire de ansamblu asupra ultimelor noutăți în domeniul atacurilor și defensivelor adversariale în contextul rețelelor neurale. Robustețea modelelor de învățare profundă este un subiect de interes intens care a atras tot mai multă atenție din partea comunității de cercetare, deoarece reprezintă un aspect important de luat în considerare în dezvoltarea și integrarea viitoarelor soluții bazate pe inteligență artificială de încredere în aplicații din viața reală. Următoarele capitole prezintă noi contribuții în acest domeniu.

Capitolul 3

Recunoașterea automată a gesturilor bazată pe semnale EMG cu utilizarea rețelelor neurale robuste

Acest capitol introduce o abordare nouă pentru construirea unui sistem robust de recunoaștere automată a gesturilor bazat pe semnale electromiografice de suprafață (sEMG), captate la nivelul antebrațului. Contribuția principală constă în propunerea unor strategii noi de învățare constrânsă care asigură robustețea împotriva perturbațiilor adverse prin controlul constantei Lipschitz a clasificatorului. Ne concentrăm pe rețele neurale cu elemente non-negative pentru care pot fi calculate limite precise ale constantei Lipschitz, și propunem diferite constrângeri ale normei spectrale care oferă garanții de robustețe din punct de vedere teoretic. Rezultatele experimentale pe patru seturi de date disponibile public arată că se poate obține un echilibru bun în ceea ce privește precizia și performanța. Demonstrăm apoi robustețea modelelor noastre, în comparație cu clasificatoarele antrenate standard, în trei scenarii, luând în considerare atât atacuri de tip cutie albă, cât și de tip cutie neagră.

3.1 EMG și recunoașterea automată a gesturilor

sEMG înseamnă electromiografie de suprafață și reprezintă manifestarea electrică a activării neuromusculare legate de contracția mușchilor [1]. Această tehnologie poate fi folosită de persoanele cu dizabilități fizice pentru a controla dispozitive de reabilitare și asistență. EMG este de asemenea folosit în multe domenii de cercetare, inclusiv în biomecanică, controlul motor, fiziologia neuromusculară, tulburările de mișcare, controlul postural și terapia fizică [32].

3.1.1 Provocări și limitări

Gesturile constituie o modalitate universală și intuitivă de comunicare, cu potențialul de a aduce experiența IoT la un nivel diferit, mai organic [33]. Algoritmii de recunoaștere auto-

mată a gesturilor (AGR) pot fi utilizați cu succes în diverse aplicații, de la recunoașterea limbajului gestual [7], la jocuri de realitate virtuală (VR) [41].

Două probleme critice trebuie abordate în dezvoltarea algoritmilor AGR: inferența suficient de rapidă pentru a asigura o senzație în timp real pentru utilizatorul final și clasificarea precisă și robustă pentru a garanta că gestul este identificat corect indiferent de condițiile de mediu. Metodele de învățare automată au devenit principalele instrumente pentru sistemele AGR, datorită capacității lor de a rezolva o mare varietate de probleme, de la regresii simple la clasificări complexe multi-modale.

Comportamentul Lipschitz al rețelei este strâns legat de rezistența acesteia împotriva atacurilor adversative.

3.2 Soluții de robustețe în contextul rețelelor neurale non-negative

3.2.1 Formularea problemei

Model 3.2.1 Orice rețea neurală de tip *feedforward* este obținută prin cascadarea a m straturi asociate cu operatorii $(T_i)_{1\leqslant i\leqslant m}$. Astfel, rețeaua neurală poate fi exprimată ca o compoziție de operatori:

$$T = T_m \circ \cdots \circ T_1. \tag{3.1}$$

Fiecare strat $i \in \{1,\ldots,m\}$ are o intrare vectorială cu valori reale x_i de dimensiune N_{i-1} care este mapată la

$$T_i(x_i) = R_i(W_i x_i + b_i), \tag{3.2}$$

unde $W_i \in \mathbb{R}^{N_i \times N_{i-1}}$, $b_i \in \mathbb{R}^{N_i}$ sunt matricea de ponderi, respectiv parametrul de decalaj. $R_i \colon \mathbb{R}^{N_i} \to \mathbb{R}^{N_i}$ constituie un operator de activare neliniar care este aplicat pe fiecare componentă (de exemplu, ReLU sau Sigmoid).

3.2.2 Certificat de robustețe Lipschitz

Considerăm o rețea neurală T așa cum este descrisă mai sus. Fie $x \in \mathbb{R}^{N_0}$ intrarea rețelei și fie $T(x) \in \mathbb{R}^{N_m}$ ieșirea asociată. Prin adăugarea unei perturbații mici $z \in \mathbb{R}^0$ la intrare, varianta ei perturbată perturbată este

$$\tilde{x} = x + z.$$

Efectul perturbației asupra ieșirii sistemului poate fi cuantificat prin următoarea inegalitate:

$$\|T(\tilde{x}) - T(x)\| \leqslant \theta_m \|z\|, \tag{3.3}$$

unde $\theta_m \geqslant 0$ reprezintă o constantă Lipschitz a rețelei. Astfel, θ_m reprezintă un parametru important care ne permite să evaluăm și să controlăm sensibilitatea unei rețele neurale la diverse perturbații. Cu toate acestea, este necesar să fie estimat cu precizie pentru a

furniza informații valoroase. O aproximare standard a constantei Lipschitz [16] este dată de

$$\theta_m = \prod_{i=1}^{m} \|W_i\|_{\mathrm{S}}, \tag{3.4}$$

unde $\|\cdot\|_{\mathrm{S}}$ denotă *norma spectrală* a unei matrice. Cu toate că este simplu de calculat, această limită aproximativă este prea largă. În literatura recentă au fost prezentate diferite metode pentru obținerea unor estimate mai strânse ale constantei Lipschitz; vezi, de exemplu, [36, 11, 13, 24, 5]. Estimările locale ale constantei Lipschitz pot, de asemenea, să fie efectuate, ceea ce poate părea mai relevant. Dar acestea sunt mai complexe de calculat și, așa cum vom vedea, controlarea constantei Lipschitz globale este, de obicei, suficientă pentru a obține o performanță bună. Estimarea constantei Lipschitz globale a rețelei este o problemă dificilă ce nu se poate rezolva în timp polinomial [36].

3.3 Metode de optimizare pentru antrenarea rețelelor neurale robuste

Pentru a asigura robustețea, vom impune constrângeri de normă spectrală asupra matricelor de ponderi. Mai precis, vom implementa o tehnică de optimizare *gradient stocastic cu proiecții*. În acest algoritm, se introduce un parametru de impuls pentru a accelera procesul de convergență.

Algorithm 1: Algoritmul de Proiecție SGD

Partition $\{1,\dots,K\}$ *în mini-loturi* $(\mathbb{L}_{q,n})_{1\leqslant q\leqslant Q}$
foreach $q \in \{1,\dots,Q\}$ **do**
  foreach $i \in \{1,\dots,m\}$ **do**
    $\Delta_{i,n} = (1+\zeta_n)\eta_{i,n} - \zeta_n \eta_{i,n-1}$ $\widetilde{\eta}_{i,n} = [(\eta_{j,n+1}^\top)_{j<i}\ \Delta_{i,n}^\top\ (\eta_{j,n}^\top)_{j>i}]^\top$
    $\eta_{i,n+1} = \mathsf{P}_{\mathscr{S}_{i,n}}\Big(\Delta_{i,n} - \gamma_n \sum_{k\in\mathbb{L}_{q,n}} \nabla_i \ell(z_k, \widetilde{\eta}_{i,n})\Big)$

unde $\mathscr{S}_{i,n} = \big\{\eta_i \mid [(\eta_{j,n+1}^\top)_{j<i}\ \eta_i^\top\ (\eta_{j,n}^\top)_{j>i}]^\top \in \mathscr{S}\big\}$.

3.3.1 Constrângeri seturi

După cum am menționat anterior, această teză se învârte în jurul rețelelor cu ponderi pozitive. Prin urmare, prima condiție pe care o impunem este non-negativitatea pentru fiecare strat $i \in \{1,\dots,m\}$, care este modelată de setul de constrângere

$$\mathscr{D}_i = \{W_i \in \mathbb{R}^{N_i \times N_{i-1}} \mid W_i \geqslant 0\} \tag{3.5}$$

Mai mult, trebuie să impunem o constrângere de normă spectrală asupra matricelor de ponderi pentru a controla robustețea sistemului. Acest lucru se traduce matematic printr-o constrângere de limită superioară:

$$\|W_m \cdots W_1\|_{\mathrm{S}} \leqslant \overline{\vartheta}, \tag{3.6}$$

unde $\overline{\vartheta}$ reprezintă limita maximă țintă a constantei Lipschitz a rețelei. Această limită constituie o măsură directă a nivelului de robustețe al sistemului față de intrări adversare. Trebuie să gestionăm aceste două constrângeri simultan pe parcursul procesului de antrnare. Pentru cea de-a doua, introducem următorul set de constrângeri.

$$\mathscr{C}_{i,n} = \{W_i \in \mathbb{R}^{N_i \times N_{i-1}} \mid \|A_{i,n} W_i B_{i,n}\|_S \leqslant \overline{\vartheta}\} \tag{3.7}$$

Astfel, obiectivul nostru va fi să efectuăm proiecția pe setul $\mathscr{S}_{i,n} = \mathscr{D}_i \cap \mathscr{C}_{i,n}$, pentru fiecare strat $i \in \{1, \ldots, m\}$ și la fiecare iterație n. Pot fi gândiți mai multi algoritmi pentru rezolvarea acestei probleme de optimizare convexe.

3.4 Configurație experimentală pentru recunoașterea gesturilor automate bazată pe sEMG

3.4.1 Seturi de date sEMG

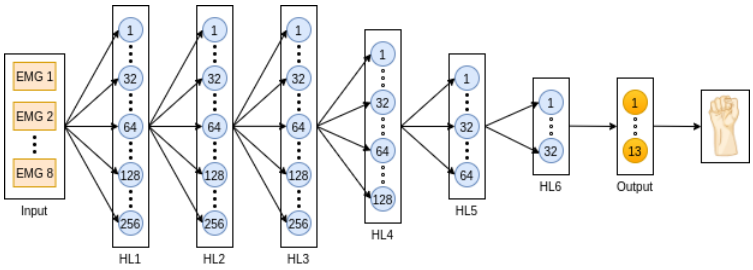


Figura 3.1 *Arhitectura propusă a rețelei neurale pentru AGR.*

Testăm schema noastră de antrenare propusă pe patru seturi de date online care conțin informații EMG despre diferite gesturi ale mâinii. Primele trei au fost achiziționate folosind brățara Myo, un dispozitiv dezvoltat de Thalmic Labs, echipat cu opt senzori sEMG amplasați circular, în timp ce ultimul a fost dobândit folosind 10 electrozi sEMG activi de tip `OttoBockMy-oBock13E200`.

Validăm, de asemenea, modelele într-un scenariu de context real. Pentru predicțiile din viața reală, am înregistrat activitatea EMG asociată fiecărui gest la nivelul antebrațului folosind brățara Myo.

3.4.2 Arhitectura propusă

Arhitectura propusă este descrisă în Figura 3.1. Semnalul brut EMG cu 8/10 canale este divizat folosind o fereastră glisantă de 250 ms, cu o suprapunere de 50%. O fereastră de 250 ms este suficient de lungă pentru a acoperi cele mai comune durate ale gesturilor, asigurând că aspectele temporale esențiale ale fiecărui gest sunt capturate în această fereastră. Suprapunerea asigură că caracteristicile semnalului importante, cum ar fi schimbările bruște sau modelele tranzitorii, nu sunt ratate din cauza frontierelor ferestrei. Din fiecare fereastră a fiecărui canal, sunt extrași opt descriptori temporali. Informația de la toate canalele este apoi concatenată, formând un vector cu 64 (80 pentru cel de-al

patrulea set de date).Informațiile provenite de la toate canalele sunt apoi concatenate, formând un vector dimensional de 64 (80 pentru al patrulea set de date).fa

3.4.3 Analiza performanței în termeni de acuratețe și robustețe

Tabel 3.1 Constanta Lipschitz obținută cu diferite strategii de constrângere pentru diferite valori ale acurateții

		Acuratețe	75%	80%	85%	90%	95%
Constanta Lipschitz 7-gestures Myo-sEMG	$\widetilde{\mathscr{C}_i} \cap \mathscr{D}_i$	$\widetilde{\mathrm{P}}_{\widetilde{\mathscr{C}_i} \cap \mathscr{D}_i}$	19.5	37.5	68.3	3.5×10^4	3.5×10^8
		$\mathrm{P}_{\widetilde{\mathscr{C}_i} \cap \mathscr{D}_i}$	0.66	13.47	74.16	1.04×10^3	1.39×10^5
	$\check{\mathscr{C}}_{i,n} \cap \mathscr{D}_i$	$\widetilde{\mathrm{P}}_{\check{\mathscr{C}}_{i,n} \cap \mathscr{D}_i}$	0.71	1.84	3.42	6.87	11.60
		$\mathrm{P}_{\check{\mathscr{C}}_i \cap \mathscr{D}_i}$	0.70	1.35	3.41	6.79	11.20
	$\mathscr{C}_{i,n} \cap \mathscr{D}_i$	$\widetilde{\mathrm{P}}_{\mathscr{C}_{i,n} \cap \mathscr{D}_i}$	0.44	1.79	2.93	4.85	5.68
		$\mathrm{P}_{\mathscr{C}_{i,n} \cap \mathscr{D}_i}$	0.35	0.46	0.65	0.82	0.95
Constanta Lipschitz 13-gestures 13Myo-sEMG	$\widetilde{\mathscr{C}_i} \cap \mathscr{D}_i$	$\widetilde{\mathrm{P}}_{\widetilde{\mathscr{C}_i} \cap \mathscr{D}_i}$	20.2	41.8	145.2	2.2×10^5	1.21×10^{11}
		$\mathrm{P}_{\widetilde{\mathscr{C}_i} \cap \mathscr{D}_i}$	0.85	20.47	112.3	1.62×10^4	2.31×10^8
	$\check{\mathscr{C}}_{i,n} \cap \mathscr{D}_i$	$\widetilde{\mathrm{P}}_{\check{\mathscr{C}}_{i,n} \cap \mathscr{D}_i}$	0.84	2.08	4.23	7.54	12.02
		$\mathrm{P}_{\check{\mathscr{C}}_i \cap \mathscr{D}_i}$	0.81	2.01	4.12	7.50	11.92
	$\mathscr{C}_{i,n} \cap \mathscr{D}_i$	$\widetilde{\mathrm{P}}_{\mathscr{C}_{i,n} \cap \mathscr{D}_i}$	0.54	1.87	3.38	4.20	5.78
		$\mathrm{P}_{\mathscr{C}_{i,n} \cap \mathscr{D}_i}$	0.49	0.53	0.75	0.92	1.25
		Acuratețe	65%	70%	75%	80%	85%
Constanta Lipschitz 24-gestures NinaPro DB5 Ex C.	$\widetilde{\mathscr{C}_i} \cap \mathscr{D}_i$	$\widetilde{\mathrm{P}}_{\widetilde{\mathscr{C}_i} \cap \mathscr{D}_i}$	25.13	57.16	188.26	2.5×10^6	2.14×10^{11}
		$\mathrm{P}_{\widetilde{\mathscr{C}_i} \cap \mathscr{D}_i}$	1.85	31.12	112.3	1.82×10^4	4.63×10^8
	$\check{\mathscr{C}}_{i,n} \cap \mathscr{D}_i$	$\widetilde{\mathrm{P}}_{\check{\mathscr{C}}_{i,n} \cap \mathscr{D}_i}$	1.74	2.41	6.02	10.17	20.14
		$\mathrm{P}_{\check{\mathscr{C}}_i \cap \mathscr{D}_i}$	1.57	2.18	5.94	10.58	19.69
	$\mathscr{C}_{i,n} \cap \mathscr{D}_i$	$\widetilde{\mathrm{P}}_{\mathscr{C}_{i,n} \cap \mathscr{D}_i}$	0.88	2.05	4.28	5.74	6.84
		$\mathrm{P}_{\mathscr{C}_{i,n} \cap \mathscr{D}_i}$	0.77	0.96	1.27	1.44	1.96
Constanta Lipschitz 53-gestures NinaPro DB 1	$\widetilde{\mathscr{C}_i} \cap \mathscr{D}_i$	$\widetilde{\mathrm{P}}_{\widetilde{\mathscr{C}_i} \cap \mathscr{D}_i}$	26.26	86.17	200.45	4.10×10^6	4.45×10^{11}
		$\mathrm{P}_{\widetilde{\mathscr{C}_i} \cap \mathscr{D}_i}$	2.60	50.12	163.14	2.8×10^4	2.9×10^9
	$\check{\mathscr{C}}_{i,n} \cap \mathscr{D}_i$	$\widetilde{\mathrm{P}}_{\check{\mathscr{C}}_{i,n} \cap \mathscr{D}_i}$	2.94	4.43	6.88	14.25	22.16
		$\mathrm{P}_{\check{\mathscr{C}}_i \cap \mathscr{D}_i}$	2.83	2.18	5.56	16.48	20.16
	$\mathscr{C}_{i,n} \cap \mathscr{D}_i$	$\widetilde{\mathrm{P}}_{\mathscr{C}_{i,n} \cap \mathscr{D}_i}$	1.22	1.80	6.83	7.40	8.23
		$\mathrm{P}_{\mathscr{C}_{i,n} \cap \mathscr{D}_i}$	1.56	2.08	2.53	2.74	3.88

Rezultatele obținute sunt rezumate în Tabelul 3.1.

3.5 Validarea Robusteții

În această secțiune, investigăm în ce măsură conceptele teoretice descrise în secțiunile anterioare ajută la îmbunătățirea robusteții clasificatorului în diferite contexte. În acest scop, luăm în considerare următoarele trei scenarii. În primul, examinăm impactul atacurilor adversariale asupra performanței clasificatorului. Al doilea scenariu ia în considerare efectul zgomotului în procesul de achiziție. În cazul semnalelor sEMG, acest zgomot poate proveni de la contactul imperfect piele-senzor cauzat de firele de păr sau picături de transpirație. În ultimul scenariu, efectuăm un experiment în viața reală folosind 10 voluntari cu capacități fizice normale.

3.5.1 Sensibilitate la Atacuri Adversariale

Evaluăm modelul nostru robust pe perturbații concepute intenționat, studiind influența acestora asupra performanței generale a sistemului. Lansăm atacuri asupra celui mai robust model în ceea ce privește acuratețea și robustețea, obținând o precizie de 92,95% și o constantă Lipschitz $\overline{\vartheta} = 0,87$ pentru setul de date cu 7 gesturi. Comparăm rezultatele cu două modele antrenate convențional: cel mai bun în ceea ce privește performanța, care atinge o acuratețe de predicție de 99,78

Pentru a crea exemplele adversariale, am folosit câteva dintre cei mai populare atacuri de tip cutie-albă, și anume: *Fast Gradient Sign Method (FGSM)* [16], *Jacobian Saliency Map Attacker (JSMA)* [30], *Projected Gradient Descent (PGD)* [28], *Carlini și Wagner (C&W)* [4], și *Gradient Matching (GM)* [15].

3.5.2 Comportament în Prezența Zgomotului

Pentru a simula efectul zgomotului generat în timpul procesului de achiziție, am adăugat zgomot sintetic direct la datele brute sEMG, înainte de etapa de extragere a caracteristicilor. Zgomotul este ales independent și distribuit identic conform unei legi mixte gaussiene $(1-p)\mathcal{N}(0,\sigma_0^2) + p\mathcal{N}(0,\sigma_1^2)$. Acest experiment evidențiază faptul că controlul constantei Lipschitz a unei rețele îmbunătățește robustețea sa nu doar împotriva atacurilor adversariale țintite, cum s-a arătat anterior, ci și în cazul atacurilor de tip cutie-neagră, unde nu se folosește nicio informație prealabilă despre model.

3.5.3 Validare într-un Scenariu din Viața Reală

Pentru a ilustra aplicabilitatea practică a descoperirilor noastre, procedăm să validăm modelul într-un context real. În acest scop, am proiectat un experiment pentru a compara un model antrenat convențional cu cel constrâns. Am observat că antrenarea unei rețele neurale pozitive supusă constrângerilor Lipschitz îmbunătățește robustețea globală a clasificatorului împotriva perturbațiilor adversariale, nu doar dintr-o perspectivă teoretică, ci și practic, generând sisteme mai fiabile cu o putere mai mare de generalizare.

3.5.4 Limitări

Una dintre principalele limite ale abordării noastre propuse este timpul de antrenare crescut. De fapt, pentru a calcula proiecția reală, metoda propusă folosește un algoritm iterativ care efectuează descompunerea valorilor singulare la fiecare iterație, operație costisitoare în resurse, în special atunci când este realizată pe matrice mari. Propunem mai multe soluții cu complexitate mai mică, care au dovedit că oferă un echilibru bun între timpul de antrenare, robustețe și performanță.

3.6 Concluzii

Acest capitol a arătat utilitatea proiectării rețelelor neurale robuste pentru recunoașterea automată a gesturilor bazate pe semnalele fiziologice sEMG. Mai precis, am propus să

controlăm cu atenție constanta Lipschitz a acestor sisteme neliniare prin considerarea arhitecturilor neurale cu ponderi pozitive. Pentru a oferi certificate de robustețe, am dezvoltat, de asemenea, noi tehnici de optimizare pentru antrenarea clasificatoarelor supuse constrângerilor asupra normei spectrale a ponderilor. Am studiat diverse formulări constrânse și am arătat că robustețea poate fi asigurată fără a sacrifica acuratețea atunci când se folosește o combinație de constrângeri strânse și proiecții exacte. De asemenea, am furnizat mai multe soluții cu o complexitate redusă, care reduc semnificativ timpul de antrenare.

Capitolul 4

Eliminarea zgomotului din semnale folosind noi clase de rețele neurale robuste

În acest capitol, ne concentrăm pe soluții robuste pentru o problemă de regresie, respectiv eliminarea zgomotului din semnalele audio. Abordăm sarcina din două perspective. În primul rând, ne concentrăm doar pe eliminarea zgomotului din elementele de magnitudine rezultate dintr-o analiză Fourier a semnalului audio. În acest scop, proiectăm o rețea complet conectată, numită Rețeaua Neurală Convoluțională Adaptabilă (ACNN), ale cărei straturi au o structură specială care prezintă o anumită similaritate cu o rețea neuronală convoluțională 1D. În a doua parte a capitolului, extindem abordarea noastră pentru a elimina zgomotul din întregul spectru complex al semnalelor audio, folosind rețele neurale cu valori complexe (CVNN). Pentru ambele soluții, derivăm limite strânse ale constantei Lipschitz și propunem mecanisme robuste de antrenare care sunt ulterior validate pe clipuri muzicale de pian, afectate de diferite niveluri de zgomot aditiv alb.

4.1 Rețele convoluționale adaptive

Acestă secțiune introduce o nouă clasă de rețele neurale, numită Rețelele Neurale Convoluționale Adaptive (ACNN), care pot fi considerate un intermediar între Rețele Neurale Convoluționale (CNN-uri) și Rețelele Complet Conectate (FCN-uri). Capacitățile de învățare ale CNN-urilor fiind bine cercetate și dovedite, profităm de acest potențial prin structurarea ponderilor rețelei noastre într-un mod similar. O diferență semnificativă constă în faptul că rețeaua folosește filtre care nu mai sunt invariante în timp/spațiu, asemănătoare cu ceea ce se face în filtrarea adaptivă.

4.1.1 Construind puntea între CNN-uri și FCN-uri

În această secțiune ne propunem să umplem golul dintre FCN-uri și CNN-uri. În termeni de concepte de prelucrare a semnalului, un strat convolutiv este un filtru multi-intrare, multi-ieșire (MIMO). Pentru semnale unidimensionale, fiecare dintre aceste filtre poate fi

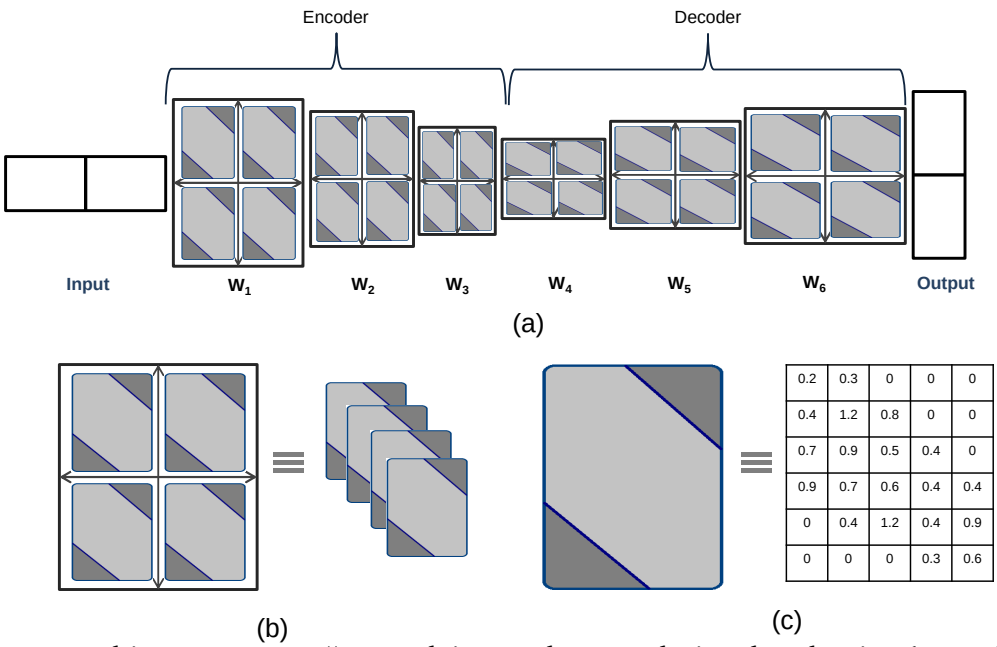


Figura 4.1 Arhitectura propusă a Rețelei Neurale Convoluționale Adaptive (ACNN). a) O arhitectură encoder-decoder compusă dintr-o rețea FCN cu 6 straturi urmată de o funcție de activare ReLU. b) Relația dintre FCN-urile propuse și CNN-uri; ponderile sunt împărțite în submatrici care simulează filtre convolutive din CNN-uri c) Fiecare dintre submatrici este restrânsă să aibă o structură de bandă, așa cum este arătat în acest exemplu. Zona de culoare gri închis marchează intrările zero, în timp ce culoarea gri deschisă corespunde celor permise să nu fie zero.

văzut ca o matrice Tœplitz generată de răspunsul la impuls al filtrului, care este aplicat la vectorul de eșantioane al semnalului. Dacă lungimea filtrului este scurtă, părțile triunghiulare superioară și inferioară ale acestei matrice sunt nule. În abordarea noastră propusă, vom păstra această structură de bandă pentru matricea de ponderi, ceea ce este echivalent cu efectuarea unei prelucrări locale la fiecare moment într-o fereastră glisantă. Cu toate acestea, pentru a adăuga mai multă flexibilitate la această arhitectură, vom permite tuturor coeficienților nenuli ai acestei matrice să fie complet optimizați. Arhitectura propusă este ilustrată în Figura 4.1a.

4.1.2 Algoritm de învățare

Pentru antrenarea ACNN propusă, folosim o optimizare de tip gradient stocastic bazată pe metoda populară ADAM [21]. Considerăm vectorul de parametri al rețelei, $\eta = (\eta_i)_{1 \leqslant i \leqslant m}$, astfel încât, pentru fiecare strat $i \in \{1, \ldots, m\}$, η_i reprezintă un vector de dimensiune $N_i(N_{i-1}+1)$, compus din elementele matricei de ponderi W_i și componente ale vectorului de decalaj b_i.

Pentru a asigura condițiile de robustețe în timp ce impunem structura dorită pentru rețeaua noastră, vectorul de parametri η este proiectat într-un set închis $\mathscr{S}$ care exprimă toate aceste constrângeri. Actualizarea parametrilor la epoca $n > 0$ este efectuată pentru loturi $(\mathbb{M}_{q,n})_{1 \leqslant q \leqslant Q}$. Dacă datele de antrenare sunt notate cu $(z_k)_{1 \leqslant k \leqslant K}$, unde z_k reprezintă

Algorithm 2: Algoritmul ADAM Proiectat

Partition $\{1,\dots,K\}$ *în mini-loturi* $(\mathbb{M}_{q,n})_{1\leqslant q\leqslant Q}$
foreach $q \in \{1,\dots,Q\}$ **do**
  $t = (n-1)Q+q$
  foreach $i \in \{1,\dots,m\}$ **do**
    $g_{i,t} = \sum_{k\in\mathbb{M}_{q,n}} \nabla_i \ell\big(z_k, (\eta_{i,t})_{1\leqslant i\leqslant m}\big)$
    $\mu_{i,t} = \beta_1\mu_{i,t-1} + (1-\beta_1)g_{i,t}$
    $\nu_{i,t} = \beta_2\nu_{i,t-1} + (1-\beta_2)g_{i,t}^2$
    $\gamma_t = \gamma\sqrt{1-\beta_2^t}/(1-\beta_1^t)$
    $\eta_{i,t+1} = \mathsf{P}_{\mathscr{S}_{i,t}}\Big(\eta_{i,t} - \gamma_t\mu_{i,t}/(\sqrt{\nu_{i,t}}+\varepsilon)\Big)$,

			PSNR	**MSE**	**CC**
Semnal cu zgomot			18.25	1.18×10^{-2}	0.76
Semnal fără zgomot	Standard - Wavelet		20.66	$1.00\times10-3$	0.80
	ACNN denoiser	Scenariu (i) $\overline{\vartheta}=1$	24.27	3.73×10^{-3}	0.96
		Scenariu (i) $\overline{\vartheta}=5$	29.03	1.25×10^{-3}	0.97
		Scenariu (i) $\overline{\vartheta}=10$	33.76	6.53×10^{-4}	0.98
		Scenariu (ii) $\overline{\vartheta}=1$	25.87	3.12×10^{-3}	0.96
		Scenariu (ii) $\overline{\vartheta}=5$	30.63	8.63×10^{-4}	0.98
		Scenariu (ii) $\overline{\vartheta}=10$	36.02	2.23×10^{-4}	0.99
	Standard FCN denoiser	$\overline{\vartheta}=1$	23.38	4.59×10^{-3}	0.90

Tabel 4.1 Compararea diferitelor variante ale metodei propuse cu metode de referință.

al k-lea cuplu de intrări și ieșiri asociate lor, operațiile efectuate în timpul epocii n sunt rezumate în Algoritmul 2, unde ridicarea la pătrat, radicalul pătrat și împărțirea sunt efectuate componentă cu componentă, și

$$\mathscr{S}_{i,t} = \big\{\eta_i \mid [(\eta_{j,t+1}^\top)_{j<i}\ \ \eta_i^\top\ \ (\eta_{j,t}^\top)_{j>i}]^\top \in \mathscr{S}\big\}. \tag{4.1}$$

4.1.3 Evaluare experimentală

Modelul propus a fost evaluat pentru eliminarea zgomotului din semnalele muzicale.

Descriere set de date

Antrenăm rețeaua ACNN propusă pe un set de date format din exerciții muzicale și melodii interpretate pe o orgă Ronald. Orga acoperă 5 octave (intervalul `C2-C7`), fiecare octavă având 12 semitonuri, generând astfel un total de 61 de note posibile. În total, setul de date conține 100 de înregistrări MIDI, cu o frecvență de eșantionare $F_s = 44100$ Hz, reprezentând 1 oră și 17 minute de înregistrare audio. Setul de date este disponibil online[1].

[1]https://speed.pub.ro/downloads/

Configurație experimentală

Datele zgomotoase destinate antrenării, validării și testării sunt generate prin adăugarea de zgomot gaussian alb la semnalele originale. Zgomotul are medie zero și deviația sa standard este aleatoare astfel încât raportul semnal-zgomot rezultat (SNR) variază între 5 și 30 dB. Eșantioanele din setul de date sunt normalizate în intervalul $[0,1]$. Extragem caracteristicile de frecvență din semnalul audio folosind o *Transformare Fourier de scurt timp* (STFT). Rețeaua estimează coeficienții STFT ai eșantioanelor, iar apoi se efectuează o *Transformare Fourier de scurt timp inversă (ISTFT)* ca pas de post-procesare. Considerăm o fereastră de analiză culisantă de tip `Hanning` de lungime $T = 23$ ms, cu o suprapunere între două ferestre consecutive de 50%. STFT este efectuat pe 1024 de puncte. În total, din fiecare segment audio se obține un vector de lungime $L = 513$ de coeficienți de frecvență, constituind intrarea în ACNN-ul nostru.

Eliminarea zgomotului este realizată folosind o arhitectură ACNN cu 6 straturi, așa cum este prezentat în Figura 4.1.

Simulări și rezultate

Pentru a măsura performanța arhitecturii ACNN propuse, efectuăm două seturi de experimente. În primul set, controlăm constanta Lipschitz a arhitecturii pentru trei valori $\overline{\vartheta}$ egale cu 1, 5 și 10. În al doilea experiment, testăm arhitectura noastră variind numărul de canale, adică modul în care divizăm fiecare matrice de ponderi.

Evaluăm performanța folosind 3 metrici standard: *Raportul maxim semnal-zgomot (PSNR)*, *Eroarea medie pătratică (MSE)*, și *Corelația încrucișată (CC)*, așa cum este arătat în Tabelul 4.1.

4.2 Proiectarea rețelelor robuste de tip feed-forward în domeniul complex

În această secțiune, introducem o nouă clasă de rețele neurale care operează în domeniul complex, numită *Robust Complex Feed-Forward Network (RCFF-Net)*. Structura rețelei este inspirată de CapsNets [6, 35].

4.2.1 Context teoretic

O rețea neurală complexă este definită astfel.

Model 4.2.1 Fie $m \in \mathbb{N} \setminus \{0\}$. T este o rețea neurală complexă cu m straturi dacă există $(N_i)_{0 \leqslant i \leqslant m} \in (\mathbb{N} \setminus \{0\})^{m+1}$ astfel încât

$$T = T_m \circ \cdots \circ T_1 \tag{4.2}$$

unde, pentru fiecare $i \in \{1, \ldots, m\}$, $T_i = R_i(W_i \cdot + b_i)$, $W_i \in \mathbb{C}^{N_i \times N_{i-1}}$, $b_i \in \mathbb{C}^{N_i}$, și $R_i \colon \mathbb{C}^{N_i} \to \mathbb{C}^{N_i}$.

În continuare, vom face ipoteza că operatorii de activare $(R_i)_{1\leqslant i\leqslant m}$ satisfac anumite proprietăți de non-expansitivate și că toți aceștia, cu excepția posibilă a ultimului strat, sunt separabili.

4.2.2 Funcții de activare complexe non-expansive

Există două rețete principale pentru construirea funcțiilor de activare, care să satisfacă condițiile impuse. Prima constă în utilizarea funcțiilor de activare *split-complex* de forma

$$(\forall z \in \mathbb{C}) \quad \rho_{i,k}(\zeta) = \rho_{i,k}^{\mathrm{R}}(\mathrm{Re}\zeta) + \imath\rho_{i,k}^{\mathrm{I}}(\mathrm{Im}\zeta) \tag{4.3}$$

unde $\rho_{i,k}^{\mathrm{R}}\colon \mathbb{R} \to \mathbb{R}$ și $\rho_{i,k}^{\mathrm{I}}\colon \mathbb{R} \to \mathbb{R}$ sunt funcții de activare de medie α_i. A doua rețetă pe care o propunem se bazează pe proprietatea următoare.

Proposition 4.2.2 *Fie $\omega\colon [0,+\infty[\to \mathbb{R}$ o funcție mediată α cu $\alpha \in]0,1]$ și astfel încât $\omega(0) = 0$. Fie ρ definită ca*

$$(\forall \zeta \in \mathbb{C}) \quad \rho(\zeta) = \begin{cases} \dfrac{\omega(|\zeta|)}{|\zeta|}\zeta & \text{dacă } \zeta \neq 0 \\ 0 & \text{altfel.} \end{cases} \tag{4.4}$$

Rezultate de robustețe

Proposition 4.2.3 *Considerați Modelul 4.2.1. Pentru fiecare $i \in \{1,\dots,m\}$, fie $W_i^+ \in [0,+\infty[^{N_i \times N_{i-1}}$. Fie $(\beta_{1,k})_{1\leqslant k\leqslant N_0} \in [0,2\pi[^{N_0}$, fie $(\beta_{m,k})_{1\leqslant k\leqslant N_m} \in [0,2\pi[^{N_1}$, și pentru fiecare $i \in \{2,\dots,m-1\}$, fie $\beta_i \in [0,2\pi[$. Presupunem că operatorii de ponderi ai rețelei sunt astfel încât*

$$\begin{aligned} &W_1 = W_1^+ \,\mathrm{Diag}\left(e^{\imath\beta_{1,1}},\dots,e^{\imath\beta_{1,N_0}}\right) \\ &(\forall i \in \{2,\dots,m-1\}) \quad W_i = e^{\imath\beta_i} W_i^+ \\ &W_m = \mathrm{Diag}\left(e^{\imath\beta_{m,1}},\dots,e^{\imath\beta_{m,N_m}}\right) W_m^+. \end{aligned} \tag{4.5}$$

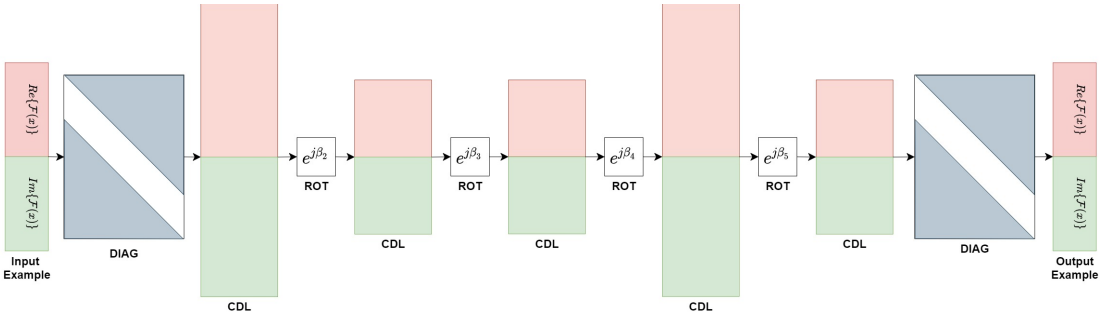
Atunci

$$\theta_m = \|W_m^+ \cdots W_1^+\|. \tag{4.6}$$

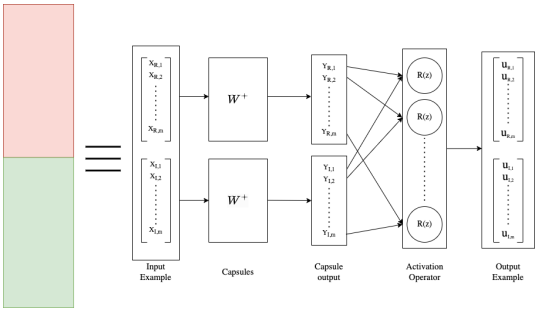
4.2.3 Abordare propusă

Rezultate experimentale pentru eliminarea zgomotului din semnale audio cu intrări atacate

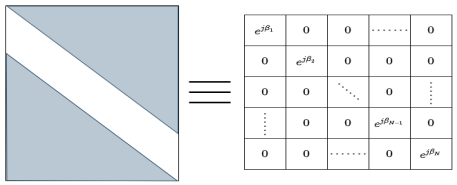
Implementăm arhitectura noastră pentru a satisface cerințele Propoziției 4.2.3 și proiectăm o Rețea Neurală Complexă (RCFF-Net). Arhitectura este ilustrată în Figura 4.2. Rețeaua procesează date complexe prin îmbinarea părților lor reale și imaginate.



(a) Arhitectura propusă: 5 CDL-uri (cu 1024, 512, 512, 1024 și 513 neuroni, respectiv) urmate de un strat de rotație (ROT) sau un strat diagonal (DIAG).



(b) Structura stratului dens complex (CDL): fiecare grup de neuroni (capsulă) va procesa în mod comun partea reală și partea imaginară a coeficienților.



(c) Structura unui strat diagonal: banda albă corespunde diagonalei principale care prezintă coeficienți nenuli.

Figura 4.2 Prezentare generală a RCFF-Network. Partea roșie denotă partea reală, în timp ce partea verde reprezintă partea imaginară.

Strategie de antrenare

În ceea ce privește strategia de antrenare, propunem să utilizăm o abordare similară cu cea a rețelelor ACNN. Folosim o versiune proiectată a optimizatorului AdaMax [21].

Algorithm 3: Algoritmul AdaMax Proiectat

Partition $\{1,\dots,K\}$ *în mini-batch-uri* $(\mathbb{M}_{q,n})_{1\leqslant q\leqslant Q}$

foreach $q \in \{1,\dots,Q\}$ **do**

$t = (n-1)Q+q$

foreach $i \in \{1,\dots,m\}$ **do**

$g_{i,t} = \sum_{k\in\mathbb{M}_{q,n}} \nabla_i \ell\big(z_k, (\eta_{i,t})_{1\leqslant i\leqslant m}\big)$

$\mu_{i,t} = \chi_1 \mu_{i,t-1} + (1-\chi_1) g_{i,t}$

$\nu_{i,t} = \max(\chi_2 \nu_{i,t-1}, |g_{i,t}|)$

$\gamma_{i,t} = \gamma \mu_{i,t}/(1-\chi_1^t)$

$\eta_{i,t+1} = \mathsf{P}_{\mathscr{S}_{i,t}}\Big(\eta_{i,t} - \gamma_t \mu_{i,t}/(\sqrt{\nu_{i,t}}+\varepsilon)\Big)$, $\eta_{i,t+1} = \mathsf{P}_{\mathscr{S}_{i,t}}(\eta_{i,t} - \gamma_{i,t}/\nu_{i,t})$

În acest algoritm, modulul și împărțirea sunt realizate pe componente. Mai sus, ℓ denotă funcția de cost, iar ∇_i reprezintă gradienții față de η_i. Vectorii $\mu_{i,t}$ și $\nu_{i,t}$ reprezintă estimările primului și celui de-al doilea moment la iterația t, utilizând parametrii $\chi_1 = 0.9$ și $\chi_2 = 0.999$. Aceste variabile sunt inițializate cu $\mu_{i,0} = \nu_{i,0} = 0$. Fiecare pas de gradient este urmat de o proiecție $\mathsf{P}_{\mathscr{S}_{i,t}}$ pe setul de constrângeri $\mathscr{S}_{i,t}$. Această mulțime exprimă cele două constrângeri pe care se bazează abordarea noastră.

Tabel 4.2 Rezultate experimentale pentru eliminarea zgomotului din semnale audio

				MSE	**PSNR [db]**	**CC**
Semnal zgomotos				7.21×10^{-3}	21.02	0.83
Linie de bază – Filtru Wiener				3.45×10^{-3}	24.24	0.94
Linie de bază – Filtru Adaptiv NLMS				2.52×10^{-3}	25.61	0.95
Linie de bază – Rețea total conectată (FCN) standardă				2.78×10^{-3}	26.05	0.95
RCFF	$\rho(\zeta) = \mathbb{C}\text{ReLU}(\zeta)$	U	$\theta_{\text{upp}} = 335$	0.96×10^{-3}	30.00	0.99
		C	$\theta_m = 0.99$	2.02×10^{-3}	27.64	0.96
	$\rho(\zeta) = \text{GK}(\zeta)$	U	$\theta_{\text{upp}} = 73.25$	1.04×10^{-3}	29.45	0.97
		C	$\theta_m = 0.99$	2.11×10^{-3}	27.14	0.96
	$\rho(\zeta) = \frac{8}{3\sqrt{3}} \frac{\lvert\zeta\rvert}{1+\lvert\zeta\rvert^2} \zeta$	U	$\theta_{\text{upp}} = 120$	0.96×10^{-3}	30.19	0.98
		C	$\theta_m = 0.93$	1.22×10^{-3}	29.02	0.97
	$\rho(\zeta) = \mathbb{C}\tanh(\zeta)$	U	$\theta_{\text{upp}} = 421$	1.28×10^{-3}	28.98	0.97
		C	$\theta_m = 0.99$	2.09×10^{-3}	27.41	0.96
	$\rho(\zeta) = \frac{\zeta}{\sqrt{1+\lvert\zeta\rvert^2}}$	U	$\theta_{\text{upp}} = 143$	1.90×10^{-3}	27.80	0.96
		C	$\theta_m = 0.97$	2.12×10^{-3}	26.98	0.96
	$\rho(\zeta) = \frac{\tanh(\lvert\zeta\rvert)}{\lvert\zeta\rvert}$	U	$\theta_{\text{upp}} = 98$	1.43×10^{-3}	28.60	0.97
		C	$\theta_m = 0.98$	1.93×10^{-3}	27.63	0.97
	$\rho(\zeta) = \zeta^{\uparrow}$	U	$\theta_{\text{upp}} = 187$	1.43×10^{-3}	30.21	0.98
		C	$\theta_m = 0.99$	1.09×10^{-3}	29.13	0.97
ACNN		C	$\theta_m = 1.00$	1.98×10^{-3}	26.24	0.96

Tabel 4.3 Rezultate experimentale pentru eliminarea zgomotului din semnale audio cu intrări atacate

				MSE	**PSNR [db]**	**CC**	Deg.[%]
Semnal zgomotos				7.30×10^{-3}	21.00	0.83	0.09
Linie de bază – Rețea total conectată (FCN) standardă				5.46×10^{-3}	22.87	0.90	12.24
RCFF	$\rho(\zeta) = \mathbb{C}\text{ReLU}(\zeta)$	U	$\theta_{\text{upp}} = 335$	4.84×10^{-3}	23.62	0.91	21.26
		C	$\theta_m = 0.99$	1.96×10^{-3}	25.43	0.95	7.99
	$\rho(\zeta) = \text{GK}(\zeta)$	U	$\theta_{\text{upp}} = 73.25$	5.42×10^{-3}	23.31	0.90	20.84
		C	$\theta_m = 0.99$	1.84×10^{-3}	25.72	0.95	5.23
	$\rho(\zeta) = \frac{8}{3\sqrt{3}} \frac{\lvert\zeta\rvert}{1+\lvert\zeta\rvert^2} \zeta$	U	$\theta_{\text{upp}} = 120$	5.26×10^{-3}	22.05	0.90	26.96
		C	$\theta_m = 0.93$	1.34×10^{-3}	28.68	0.97	1.17
	$\rho(\zeta) = \mathbb{C}\tanh(\zeta)$	U	$\theta_{\text{upp}} = 421$	5.15×10^{-3}	23.14	0.90	22.14
		C	$\theta_m = 0.99$	2.82×10^{-3}	25.41	0.95	6.20
	$\rho(\zeta) = \frac{\zeta}{\sqrt{1+\lvert\zeta\rvert^2}}$	U	$\theta_{\text{upp}} = 143$	6.02×10^{-3}	22.24	0.89	26.45
		C	$\theta_m = 0.97$	2.98×10^{-3}	25.12	0.94	8.14
	$\rho(\zeta) = \frac{\tanh(\lvert\zeta\rvert)}{\lvert\zeta\rvert}$	U	$\theta_{\text{upp}} = 98$	5.78×10^{-3}	21.36	0.89	23.32
		C	$\theta_m = 0.98$	5.46×10^{-3}	25.56	0.95	5.61
	$\rho(\zeta) = \zeta^{\uparrow}$	U	$\theta_{\text{upp}} = 187$	4.67×10^{-3}	23.09	0.90	22.34
		C	$\theta_m = 0.99$	1.45×10^{-3}	28.20	0.95	2.60
ACNN		C	$\theta_m = 1.00$	2.46×10^{-3}	25.43	0.95	3.08

4.2.4 Rezultate experimentale

Metodologia propusă este aplicată aceleiași probleme ca în secțiunea anterioară. Folosim o rețea RFCC-Net cu 5 straturi ($m = 5$), cu diverse funcții de activare, și folosim aceeași prelucrare a datelor ca în Secțiunea 4.1.3.

Principala diferență este că rețeaua estimează acum coeficienții complexi STFT și, în faza de postprocesare, se realizează o operație inversă (ISTFT) pentru reconstrucția semnalului.

Evaluăm performanța RCFF-Net pe aceleași 3 metrici standard: *Raportul maxim semnal-zgomot (PSNR)*, *Corelația încrucișată (CC)*, și *Eroarea medie pătratică (MSE)*, care a fost, de asemenea, utilizată ca ca funcție de cost. Rezultatele pe setul de testare sunt rezumate în Tabelul 4.2. Comparăm soluția noastră cu alte tehnici standard de reducere a zgomotului, respectiv filtrul Wiener optim și filtrul adaptiv bazat pe algoritmul

Normalised Least Mean Squares (NLMS). Ca altă bază de comparație, am antrenat și o rețea clasică cu 5 straturi (FCN) cu activare ReLU. În plus, am antrenat RCFF-Net atât în variantă constrânsă, cât si standard denumite în Tabel 4.2 ca `C` și respectiv `U`.

4.3 Concluzie

Acest capitol propune două noi clase de rețele neurale. Prima, ACNN, stabilește o legătură nouă între straturile complet conectate și structurile convoluționale, în timp ce a doua, RCFF-Net, operează în spațiul complex. Prin structurarea judicioasă a matricelor de ponderi, am derivat o limită strânsă Lipschitz pentru ambele arhitecturi propuse. În cazul complex, analiza noastră a condus la noi rezultate teoretice privind funcțiile de activare non-expansive. Am extins, de asemenea, o limită Lipschitz strânsă existentă pentru rețelele neurale ce operează în domeniul complex. Calcularea acestei limite nu mai este o problemă combinatorială pentru rețelele neurale complexe, ceea ce evidențiază provocările ridicate în ceea ce privește cazul real. De asemenea, am arătat cum să controlăm numeric limitele Lipschitz în procesul de antrenament.

Capitolul 5

Rețele neurale ABBA: gestionarea pozitivității, expresivității și robusteții

În acest capitol, introducem rețelele ABBA, o clasă nouă de rețele neurale (aproape) non-negative, care se dovedesc fi echipate cu o serie de proprietăți atractive. În special, demonstrăm că aceste rețele sunt aproximatoare universale, beneficiind în același timp de avantajele rețelelor ponderate non-negative. Derivăm limite strânse Lipschitz atât în cazurile de rețele complet conectate, cât și în cele convoluționale. Propunem o strategie pentru proiectarea rețelelor ABBA care sunt robuste împotriva atacurilor adversariale, controlând fin constanta Lipschitz a rețelei în timpul fazei de antrenare. Demonstrăm că metoda noastră depășește alte metode de apărare de ultimă generație împotriva atacatorilor adversari de tip cutie-albă. Experimentele sunt efectuate în sarcini de clasificare a imaginilor pe patru seturi de date de referință.

5.1 Introducere

Este larg acceptat faotul că oamenii posedă abilitatea nativă de a decompune interacțiunile complexe în categorii discrete, intuitive și ierarhice, înainte de a le analiza [25]. Conceptual, această evoluție către o reprezentare bazată pe blocuri în cogniția umană poate fi legată de restricțiile non-negative asupra ponderilor rețelei [9]. Această idee, împreună cu alți factori, a stârnit interesul în rețelele neurale cu ponderi non-negative.

Abordare. Suntem interesați de rețelele neurale cu ponderi non-negative, cu excepția primului și ultimului strat liniar. Ne concentrăm asupra unei subclase particulare a acestor rețele, pentru care matricele de ponderi au o structură de forma

$$\begin{bmatrix} A & B \\ B & A \end{bmatrix},$$

beneficiind astfel de mai multe proprietăți algebrice. Rețelele corespunzătoare sunt numite ulterior rețele ABBA. De menționat că matricele de ponderi A și B sunt duplicate în rețelele ABBA, permițându-ne să limităm numărul de parametri.

5.2 Lucrări conexe

Rețele neurale non-negative. Inspirată de tehnicile de factorizare a matricelor non-negative (NMF), lucrarea [9] introduce restricții non-negative asupra ponderilor pentru a crea rețele neurale în care unitățile ascunse corespund conceptelor identificabile.

Legături cu alte rețele. Din altă perspectivă, ideea de a folosi ponderi redundante amintește de rețelele siameze [2]. Aceste arhitecturi sunt folosite cu succes pentru a gestiona sarcini de învățare a similarității, cum ar fi verificarea feței [39], recunoașterea caracterelor [22] și urmărirea obiectelor [19].

5.3 Rețele neurale ABBA

5.3.1 Formularea problemei

Vom spune că operatorul de activare R_i este simetric, dacă există $(c_i, d_i) \in (\mathbb{R}^{N_i})^2$ astfel încât

$$(\forall x \in \mathbb{R}^{N_i}) \quad R_i(x) - d_i = -R_i(-x + c_i). \tag{5.1}$$

Altfel spus, $(c_i, d_i)/2$ reprezintă un centru de simetrie al graficului lui R_i.

5.3.2 Matricele ABBA

Mai întâi definim matricele ABBA, care vor fi principala unealtă algebraică în acest capitol.

Definition 5.3.1 Fie $(N_1, N_2) \in (\mathbb{N} \setminus \{0\})^2$. $\mathscr{A}_{N_1,N_2}$ este mulțimea matricelor ABBA de dimensiune $(2N_2) \times (2N_1)$, adică $M \in \mathscr{A}_{N_1,N_2}$ dacă există matricele $A \in \mathbb{R}^{N_2 \times N_1}$ și $B \in \mathbb{R}^{N_2 \times N_1}$ astfel încât

$$M = \begin{bmatrix} A & B \\ B & A \end{bmatrix}. \tag{5.2}$$

Matricea sumă asociată cu M este apoi definită ca $\mathfrak{S}(M) = A + B$.

5.3.3 Extindere la rețelele neurale

În această secțiune prezentăm rețeaua neurală ABBA pentru straturile complet conectate.

Definition 5.3.2 Fie $m \in \mathbb{N} \setminus \{0\}$. $\widetilde{T}$ este o rețea ABBA cu m straturi dacă

$$\widetilde{T} = (\widetilde{W}_{m+1} \cdot + \widetilde{b}_{m+1}) \widetilde{T}_m \cdots \widetilde{T}_1 \widetilde{W}_0 \tag{5.3}$$

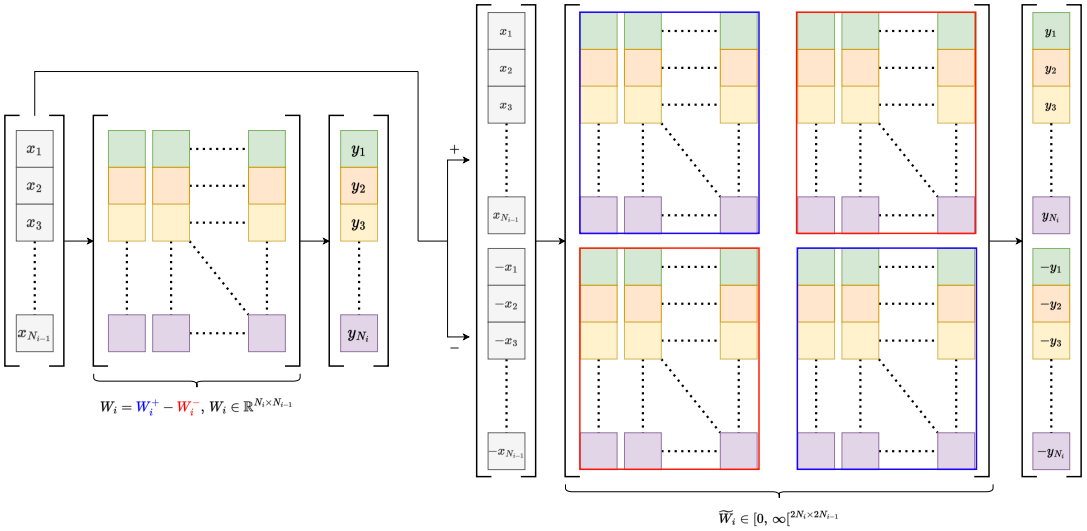


Figura 5.1 Echivalență între un strat standard complet conectat și corespondentul său ABBA.

cu $\widetilde{W}_0 \in \mathbb{R}^{(2N_0)\times N_0}$, $\widetilde{W}_{m+1} \in \mathbb{R}^{N_m \times (2N_m)}$, $\widetilde{b}_{m+1} \in \mathbb{R}^{N_m}$, și

$$(\forall i \in \{1,\dots,m\}) \quad \widetilde{T}_i = \widetilde{R}_i(\widetilde{W}_i \cdot + \widetilde{b}_i) \tag{5.4}$$

$$\widetilde{R}_i \colon \mathbb{R}^{2N_i} \to \mathbb{R}^{2N_i}, \tag{5.5}$$

$$\widetilde{b}_i \in \mathbb{R}^{2N_i}, \tag{5.6}$$

$$\widetilde{W}_i \in \mathscr{A}_{N_i,N_{i-1}}, \tag{5.7}$$

pentru numerele întregi pozitive date $(N_i)_{0\leqslant i\leqslant m}$. $\widetilde{T}$ este o rețea ABBA cu m straturi non-negative dacă este o rețea ABBA cu m straturi așa cum este definită mai sus și, pentru fiecare $i \in \{1,\dots,m\}$, elementele din $\widetilde{W}_i$ sunt non-negative.

5.3.4 Legătura cu rețelele neurale standard

O ilustrare a legăturii dintre straturile complet conectate și matricele ABBA este prezentată în Figura 5.1.

5.3.5 Expresivitatea rețelelor ABBA non-negative

Unul dintre principalele avantaje ale rețelelor ABBA non-negative față de rețelele standard cu ponderi non-negative este că acestea sunt aproximatoare universale.

5.3.6 Limite Lipschitz pentru rețelele ABBA complet conectate

În această secțiune, arătăm că putem deriva o expresie simplă pentru constanta Lipschitz, folosind o limită separabilă, pentru rețelele ABBA non-negative.

Proposition 5.3.3 *Fie $m \in \mathbb{N}\setminus\{0\}$ și fie $\widetilde{T} \in \mathscr{N}^+_{m,\mathscr{A}}$ dată de* (5.3)-(5.7). *Presupunem că, pentru fiecare $i \in \{1,\dots,m-1\}$, $\widetilde{R}_i$ este un operator separabil nonexpansiv. O constantă*

Lipschitz a lui $\widetilde{T}$ este

$$\theta_m = \|\widetilde{W}_{m+1}\| \ \|\mathfrak{S}(\widetilde{W}_m)\cdots\mathfrak{S}(\widetilde{W}_1)\| \, \|\widetilde{W}_0\|. \tag{5.8}$$

Constanta Lipschitz a lui $\widetilde{T}$ în (5.8) se reduce la

$$\theta_m = \||W_m|\ldots|W_1|\|. \tag{5.9}$$

5.4 Rețele convoluționale

Aici extindem rezultatele prezentate în Secțiunea 5.3 la straturile convoluționale.

5.4.1 Straturi convoluționale ABBA

Stratul convoluțional ABBA $\widetilde{W}_i$ are de două ori numărul de canale de intrare și de două ori numărul de canale de ieșire. În această secțiune arătăm modelarea matematică a unui strat convoluțional ABBA.

5.4.2 Limite Lipschitz pentru rețelele convoluționale

În această secțiune, stabilim limite pentru constanta Lipschitz a unei rețele neurale convoluționale T cu m straturi.

Theorem 5.4.1 *Fie $(\sigma_i)_{1\leqslant i\leqslant m}$ factorii agregati de pasi ai rețelei T, și fie*

$$W = (W_m)_{\uparrow\sigma_{m-1}} * \cdots * (W_2)_{\uparrow\sigma_1} * W_1 \tag{5.10}$$

unde $(W_i)_{1\leqslant i\leqslant m}$ sunt răspunsurile impulsive MIMO ale fiecărui strat al rețelei T și, pentru fiecare $i \in \{2,\ldots,m\}$, $(W_i)_{\uparrow\sigma_{i-1}}$ este secvența interpolată cu un factor σ_{i-1} a lui W_i. Pentru fiecare $\mathbf{j} \in \mathbb{S}(\sigma_m) = \{0,\ldots,\sigma_m - 1\}^d$, definim următoarea matrice:

$$\overline{W}^{(\mathbf{j})} = \sum_{\mathbf{n}\in\mathbb{Z}^d} W(\sigma_m\mathbf{n}+\mathbf{j}) \in [0,+\infty[^{\zeta_m\times\zeta_0}. \tag{5.11}$$

Atunci

$$\theta_m = \Big\| \sum_{\mathbf{j}\in\mathbb{S}(\sigma_m)} \overline{W}^{(\mathbf{j})} \big(\overline{W}^{(\mathbf{j})}\big)^\top \Big\|^{1/2} \tag{5.12}$$

este o limită inferioară pentru estimarea constantei Lipschitz a rețelei T. În plus, dacă pentru fiecare $i \in \{1,\ldots,m\}$, $p \in \{1,\ldots,\zeta_{i-1}\}$, și $q \in \{1,\ldots,\zeta_i\}$, $w_{i,q,p} = (w_{i,q,p}(\mathbf{n}))_{\mathbf{n}\in\mathbb{Z}^d}$ este un nucleu non-negativ, atunci θ_m este o constantă Lipschitz a lui T.

5.4.3 Limite pentru rețelele convoluționale ABBA

Aici extindem rezultatele anterioare la contextul ABBA.

Theorem 5.4.2 *Sub presupunerile anterioare referitoare la rețeaua convoluțională ABBA $\widetilde{T}$, fie*

$$(\forall i \in \{1,\dots,m\})(\forall \mathbf{j} \in \mathbb{S}(s_i)) \quad \Omega_i^{(\mathbf{j})} = \sum_{\mathbf{n}\in\mathbb{Z}^d} \mathfrak{S}(\widetilde{W}_i(s_i\mathbf{n}+\mathbf{j})) \in [0,+\infty[^{\zeta_i\times\zeta_{i-1}}, \tag{5.13}$$

unde $(\widetilde{W}_i(\mathbf{n}))_{n\in\mathbb{Z}}$ *este răspunsul impuls MIMO al stratului ABBA de index i. Atunci o constantă Lipschitz a lui* $\widetilde{T}$ *este*

$$\overline{\theta}_m = \|\widetilde{W}_{m+1}\| \Big(\prod_{i=1}^{m} \Big\| \sum_{\mathbf{j}\in\mathbb{S}(s_i)} \Omega_i^{(\mathbf{j})} \big(\Omega_i^{(\mathbf{j})}\big)^\top \Big\| \Big)^{1/2} \|\widetilde{W}_0\|, \tag{5.14}$$

unde $\|\widetilde{W}_{m+1}\|$ *(resp.* $\|\widetilde{W}_0\|$*) este norma spectrală a operatorului liniar folosit în ultimul (respectiv primul) strat.*

5.5 Mecanism de antrenare cu constrângeri Lipschitz

Algorithm 4: Algoritmul ADAM Proiectat

Partition $\{1,\dots,K\}$ *în mini-loturi* $(\mathbb{L}_{q,n})_{1\leqslant q\leqslant Q}$
$t = (n-1)Q+q$ # index de iterație
parcurgere mini-loturi
foreach $q \in \{1,\dots,Q\}$ **do**
  foreach *strat i* **do**
    $g_{i,t} = \sum_{k\in\mathbb{M}_{q,n}} \nabla_i \ell\big(z_k, (\Psi_{i,t})_{1\leqslant i\leqslant m}\big)$ # calcul gradient
    $\mu_{i,t} = \beta_1\mu_{i,t-1} + (1-\beta_1)g_{i,t}$ # actualizări clasice ADAM
    $\nu_{i,t} = \beta_2\nu_{i,t-1} + (1-\beta_2)g_{i,t}^2$
    $\gamma_t = \gamma\sqrt{1-\beta_2^t}/(1-\beta_1^t)$
    $\widetilde{\Psi}_{i,t} = \Psi_{i,t} - \gamma_t\mu_{i,t}/(\sqrt{\nu_{i,t}}+\varepsilon)$
  foreach *strat i* **do**
    $\Psi_{i,t+1} = \mathrm{proj}_{\mathscr{S}_{i,t}}(\widetilde{\Psi}_{i,t})$ # pas de proiecție

5.6 Experimente

În această secțiune, prezentăm versatilitatea rețelelor neurale ABBA în rezolvarea sarcinilor de clasificare. Obiectivul experimentelor noastre este triplu.

(i) În primul rând, comparăm structurile ABBA pozitive cu echivalentele lor clasice non-negative și verificăm că metoda noastră produce rezultate semnificativ mai bune în toate cazurile considerate.

(ii) Apoi, antrenăm modele ABBA cu constrângeri diferite ale constantei Lipschitz și evaluăm robustețea acestora împotriva mai multor atacuri adversariale.

(iii) În final, comparăm abordarea propusă cu alte trei strategii de apărare, respectiv *Antrenament Adversarial* (AT), *Trade-off-inspired adversarial defense* (TRADES) [43], și *Deel-Lip* propusă de [37].

Set de date	Rețea	Arhitectură	Precizie [%]
MNIST	ABBA	Dense	98.33
		Conv	98.70
	Non-Negative	Dense	94.95
		Conv	93.27
	standard	Dense	98.35
		Conv	98.68
FMNIST	ABBA	Dense	90.02
		Conv	90.17
	Non-Negative	Dense	84.56
		Conv	83.09
	standard	Dense	90.00
		Conv	90.20
RPS	ABBA	Conv	99.08
	Non-Negative	Conv	67.30
	standard	Conv	98.86
CelebA	ABBA	Conv	90.21
	Non-Negative	Conv	61.04
	standard	Conv	90.17

Tabel 5.1 Comparare între rețelele ABBA, complet non-negative și cu semn arbitrar (standard).

5.7 Concluzii

În acest capitol, introducem rețelele ABBA, o clasă nouă de rețele neurale în care majoritatea ponderilor sunt non-negative. Demonstrăm că aceste rețele sunt aproximatori universali, posedând toate proprietățile expresive ale arhitecturilor neurale semnate convenționale. În plus, dezvăluim caracteristicile lor algebrice remarcabile, permițându-ne să derivăm limite precise Lipschitz pentru operatorii complet conectați și convolutivi.

Exploatând aceste limite, construim rețele neurale robuste potrivite pentru diverse sarcini de clasificare. Pentru cercetările viitoare, ar fi interesant să explorăm aplicarea rețelelor ABBA în probleme de regresie, unde controlul constantei Lipschitz poate prezenta mai multe provocări. Mai mult, extinderea limitelor noastre teoretice la diferite structuri, cum ar fi rețelele recurente sau cele bazate pe atenție, promite progrese suplimentare.

În final, recunoaștem necesitatea investigării scalabilității metodei de antrenare propuse pentru arhitecturi profunde. Una dintre principalele provocări în acest demers este numărul crescut de parametri pe care arhitecturile ABBA profunde le presupun.

Capitolul 6

Concluzii

6.1 Sumar

În ciuda faptului că pot părea în fruntea dezvoltărilor în Știința Datelor, rețelele neurale generează provocări în domeniile securității, confidențialității și siguranței din cauza susceptibilității lor la o varietate largă de amenințări și perturbări care pot apărea în timpul funcționării lor. Prin urmare, este vital să înțelegem motivele instabilității rețelelor neurale, să identificăm vulnerabilitățile și să dezvoltăm soluții care să îmbunătățească stabilitatea acestora pentru a garanta existența sistemelor bazate pe IA care sunt agnostice la mici variații ale intrărilor lor.

În timpul acestei teze, accentul nostru principal a fost pe proiectarea și antrenarea rețelelor neurale care sunt intrinsec robuste împotriva perturbațiilor adversariale ale intrărilor. Astfel, am propus mai multe tehnici de antrenare robustă și am dovedit eficacitatea lor în rezolvarea problemelor de clasificare și regresie. Am arătat că cercetarea noastră este aplicabilă pe o gamă largă de aplicații și că rezultatele sale pot fi utile și în scenarii din viața reală.

În primul rând, ne-am concentrat pe rețelele de tip *feed-forward*, care conțin doar straturi liniare. Cercetarea noastră a început de la rezultatele stabilite în [11], care afirmă că în cazul rețelelor neurale ponderate pozitiv, se pot deriva limite Lipschitz strânse. Am proiectat mai mulți algoritmi de antrenare robustă, încercând să obținem un echilibru bun între robustețe și performanță.

6.2 Perspective

În această secțiune, propunem câteva extinderi posibile ale metodelor menționate anterior, care ar putea merita investigarea în viitoarele lucrări.

6.2.1 Antrenarea sistemelor de eliminare a zgmotului 1-Lipschitz

O modalitate posibilă de extindere a lucrării prezentate în această teză ar fi să folosim metodele noastre stabilite pentru controlul constantei Lipschitz a rețelelor neurale pentru a genera sisteme de eliminare de zgomot 1-Lipschitz, așa cum este prezentat în [34].

6.2.2 Extinderea aplicațiilor rețelelor neurale complexe

În viitoarele lucrări, ar fi interesant să aplicăm RCFF-Net la o gamă mai largă de aplicații de prelucrare a semnalelor care implică date complexe, cum ar fi separarea surselor audio unde rețelele robuste CVNN pot juca un rol esențial.

6.2.3 Controlul constantei Lipschitz pentru structuri de straturi mai complexe

Având în vedere progresul realizat în această teză, în special în gestionarea eficientă a constantei Lipschitz pentru a îmbunătăți stabilitatea straturilor liniare și convoluționale din rețelele neurale, apare o perspectivă captivantă pentru viitoare cercetări de a extinde aceste metode la structuri mai complexe, cum ar fi cele recurente.

6.2.4 Combinația controlului Lipschitz cu alte metode de apărare certificate

În contextul îmbunătățirii stabilității rețelelor neurale împotriva amenințărilor adversariale, o direcție promițătoare pentru cercetările viitoare constă în integrarea mecanismelor noastre actuale de control al constantei Lipschitz cu strategii de apărare complementare. Deosebit de interesantă este potențiala sinergie dintre abordarea noastră și apărările certificate, cum ar fi GloRoNets [26].

6.2.5 Studiul efectului altor tehnici de regularizare

O altă direcție interesantă ar fi studiul cuprinzător al efectelor diferitelor tehnici de regularizare asupra stabilității modelului.

6.2.6 Extinderea la alte distanțe

Extinderea metodelor noastre actuale pentru controlul robusteții rețelelor neurale pentru a acoperi alte metrici reprezintă o altă perspectivă de cercetare. În prezent, tehnicile noastre abordează în principal perturbațiile ℓ_2, dar funcționalitatea sistemelor din lumea reală impune o abordare mai cuprinzătoare [4].

Bibliografie

[1] Atzori, M., Gijsberts, A., Castellini, C., Caputo, B., Hager, A.-G. M., Elsig, S., Giatsidis, G., Bassetto, F., and Müller, H. (2014). Electromyography data for non-invasive naturally-controlled robotic hand prostheses. *Sci. data*, (140053).

[2] Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., and Shah, R. (1993). Signature verification using a "siamese" time delay neural network. In *Proc. Ann. Conf. Neur. Inform. Proc. Syst.*, volume 6, pages 737–744.

[3] Carlini, N., Mishra, P., Vaidya, T., Zhang, Y., Sherr, M., Shields, C., Wagner, D., and Zhou, W. (2016). Hidden voice commands. In *USENIX Security Symp.*, pages 513–530.

[4] Carlini, N. and Wagner, D. (2017). Towards evaluating the robustness of neural networks. In *IEEE Symp. Security and Privacy*, pages 39–57.

[5] Chen, T., Lasserre, J.-B., Magron, V., and Pauwels, E. (2020). Semialgebraic optimization for Lipschitz constants of ReLU networks. pages 19189–19200.

[6] Cheng, X., He, J., He, J., and Xu, H. (2019). Cv-CapsNet: Complex-valued Capsule Network. *IEEE Access*, 7:85492–85499.

[7] Cheok, M. J., Omar, Z., and Jaward, M. H. (2019). A review of hand gesture and sign language recognition techniques. *Int. J. Mach. Learn. Cyber.*, 10(1):131–153.

[8] Chorowski, J. and Zurada, J. M. (2015). Learning understandable neural networks with nonnegative weight constraints. *IEEE Trans. Neural Netw. Learn. Syst.*, 26(1):62–69.

[9] Chorowski, J. and Zurada, J. M. (2015). Learning understandable neural networks with nonnegative weight constraints. In *IEEE Trans. Neural Net. and Learn. Syst.*, volume 26, pages 62–69.

[10] Cisse, M., Bojanowski, P., Grave, E., Dauphin, Y., and Usunier, N. (2017). Parseval networks: Improving robustness to adversarial examples. In *Proc. Int. Conf. Mach. Learn.*, pages 854–863.

[11] Combettes, P. L. and Pesquet, J.-C. (2020). Lipschitz certificates for layered network structures driven by averaged activation operators. In *J. Math. Data Sci.*, volume 2, pages 529–557.

[12] Drucker, H. and LeCun, Y. (1992). Improving generalization performance using double backpropagation. *IEEE Trans. Neural Netw. Learn. Syst.*, 3:991–997.

[13] Fazlyab, M., Robey, A., Hassani, H., Morari, M., and Pappas, G. (2019). Efficient and accurate estimation of lipschitz constants for deep neural networks. In *Proc. Ann. Conf. Neur. Inform. Proc. Syst.*, pages 11423–11434.

[14] Gamarnik, D., Kızıldağ, E. C., and Zadik, I. (2021). Self-regularity of output weights for overparameterized two-layer neural networks. In *Proc. IEEE Int. Symp. Info. Theory*, pages 819–824.

[15] Geiping, J., Fowl, L. H., Huang, W. R., Czaja, W., Taylor, G., Moeller, M., and Goldstein, T. (2020). Witches' Brew: Industrial scale data poisoning via Gradient Matching. In *Int. Conf. Learn. Represent.*

[16] Goodfellow, I., Shlens, J., and Szegedy, C. (2015). Explaining and harnessing adversarial examples. In *Proc. Int. Conf. Learn. Represent.*

[17] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C. (2017). Improved training of wasserstein gans. In *Proc. Ann. Conf. Neur. Inform. Proc. Syst.*, pages 5767–5777.

[18] Guo, C., Karrer, B., Chaudhuri, K., and van der Maaten, L. (2022). Bounding training data reconstruction in private (deep) learning. In *Proc. Int. Conf. Machine Learn.*, pages 8056–8071.

[19] He, A., Luo, C., Tian, X., and Zeng, W. (2018). A twofold siamese network for real-time object tracking. In *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, pages 4834–4843.

[20] He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., and Zhang, K. (2019). The practical implementation of artificial intelligence technologies in medicine. *Nat. med.*, 25(1):30–36.

[21] Kingma, D. P. and Ba, J. (2015). Adam: A method for stochastic optimization. *Proc. Int. Conf. Learning Represent.*

[22] Koch, G., Zemel, R., and Salakhutdinov, R. (2015). Siamese neural networks for one-shot image recognition. In *Proc. Int. Conf. Machine Learn.*, volume 2.

[23] Kurakin, A., Goodfellow, I., and Bengio, S. (2016). Adversarial machine learning at scale. In *Proc. Int. Conf. Learn. Represent.*

[24] Latorre, F., Rolland, P., and Cevher, V. (2020). Lipschitz constant estimation of neural networks via sparse polynomial optimization. In *Int. Conf. on Learning Represent.*

[25] Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. In *Nature*, volume 401, pages 788–791.

[26] Leino, K., Wang, Z., and Fredrikson, M. (2021). Globally-robust neural networks. In *Proc. Int. Conf. Mach. Learn.*, pages 6212–6222.

[27] Li, P., Chen, X., and Shen, S. (2019). Stereo R-CNN based 3D object detection for autonomous driving. In *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, pages 7644–7652.

[28] Madry, A., Makelov, A., Schmidt, L., Tsipras, D., and Vladu, A. (2018). Towards deep learning models resistant to adversarial attacks. In *Proc. Int. Conf. Learn. Represent.*

[29] Neacşu, A., Pesquet, J.-C., and Burileanu, C. (2020). Accuracy-robustness trade-off for positively weighted neural networks. In *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, pages 8389–8393.

[30] Papernot, N., McDaniel, P., Jha, S., Fredrikson, M., Celik, Z. B., and Swami, A. (2016). The limitations of deep learning in adversarial settings. In *IEEE Symp. Security Privacy*.

[31] Parkinson, S., Ongie, G., and Willett, R. (2023). Linear neural network layers promote learning single- and multiple-index models. In *arXiv:2305.15598*.

[32] Pauk, J. (2008). Different techniques for EMG signal processing. *J. of Vibroeng.*, 10:571–576.

[33] Qi, J., Jiang, G., Li, G., Sun, Y., and Tao, B. (2019). Intelligent human-computer interaction based on surface EMG gesture recognition. *IEEE Access*, 7:61378–61387.

[34] Repetti, A., Terris, M., Wiaux, Y., and Pesquet, J.-C. (2022). Dual forward-backward unfolded network for flexible Plug-and-Play. In *Proc. European Signal Processing Conference*, pages 957–961.

[35] Sabour, S., Frosst, N., and Hinton, G. E. (2017). Dynamic routing between capsules. In *Proc. Ann. Conf. Neur. Inform. Proc. Syst.*, volume 30, pages 3856–3866.

[36] Scaman, K. and Virmaux, A. (2018). Lipschitz regularity of deep neural networks: analysis and efficient estimation. In *Proc. Ann. Conf. Neur. Inform. Proc. Syst.*, pages 3839–3848.

[37] Serrurier, M., Mamalet, F., González-Sanz, A., Boissin, T., Loubes, J.-M., and Del Barrio, E. (2021). Achieving robustness in classification using optimal transport with hinge regularization. In *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, pages 505–514.

[38] Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., and Fergus, R. (2014). Intriguing properties of neural networks. In *Proc. Int. Conf. Learn. Represent.*

[39] Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, pages 1701–1708.

[40] Takeru, M., Toshiki, K., Masanori, K., and Yuichi, Y. (2018). Spectral normalization for generative adversarial networks. In *Int. Conf. Learn. Represent.*

[41] Wen, F., Sun, Z., He, T., Shi, Q., Zhu, M., Zhang, Z., Li, L., Zhang, T., and Lee, C. (2020). Machine learning glove using self-powered conductive superhydrophobic triboelectric textile for gesture recognition in VR/AR applications. *Adv. Sci.*, 7(14):2000261.

[42] Widiastuti, N. (2019). Convolution neural network for text mining and natural language processing. In *IOP Conf.: Mater. Sci. Eng.*, volume 662.

[43] Zhang, H., Yu, Y., Jiao, J., Xing, E., El Ghaoui, L., and Jordan, M. (2019). Theoretically principled trade-off between robustness and accuracy. In *Proc. Int. Conf. Machine Learn.*, pages 7472–7482.