

REDE NEURAL PIRAMIDAL COM RESTRIÇÃO DE PESOS NÃO-NEGATIVOS PARA APRENDIZAGEM POR PARTES DE PADRÕES VISUAIS

Trabalho de Conclusão de Curso
Engenharia da Computação

Milla Silva Alcoforado Ferro

Orientador: Prof. Dr. Bruno José Torres Fernandes

Milla Silva Alcoforado Ferro

**REDE NEURAL PIRAMIDAL COM RESTRIÇÃO DE
PESOS NÃO-NEGATIVOS PARA APRENDIZAGEM
POR PARTES DE PADRÕES VISUAIS**

Monografia apresentada como requisito parcial para obtenção do diploma de Bacharel em Engenharia de Computação pela Escola Politécnica de Pernambuco - Universidade de Pernambuco

Universidade de Pernambuco
Escola Politécnica de Pernambuco
Graduação em Engenharia de Computação

Orientador: Prof. Dr. Bruno José Torres Fernandes

Recife - PE, Brasil

12 de abril de 2015

Declaro que revisei o Trabalho de Conclusão de Curso sob o título “*REDE NEURAL PIRAMIDAL COM RESTRIÇÃO DE PESOS NÃO-NEGATIVOS PARA APRENDIZAGEM POR PARTES DE PADRÕES VISUAIS*”, de autoria de *Milla Silva Alcoforado Ferro*, e que estou de acordo com a entrega do mesmo.

Recife, _____ / _____ / _____

Prof. Dr. Bruno José Torres Fernandes
Orientador

Dedico este trabalho à minha maior fonte de motivação: minha família.

Agradecimentos

Primeiramente, agradeço a Deus por ter me guiado até o fim de mais uma etapa importante de minha vida. Agradeço aos meus pais Cláudio e Mirian por me ensinarem o quão importante é investir na minha educação. Eles não mediram esforços para que eu chegasse até aqui.

Agradeço a meu esposo Douglas pela paciência e motivação de sempre.

Agradeço ao professor Bruno que não só me orientou na elaboração deste trabalho como também em toda a minha graduação. Ao professor Byron por ter me apoiado durante minha iniciação científica e compartilhado conhecimento juntamente com os membros do grupo de pesquisa RPPDI. Agradeço ao professor Gustavo pelo auxílio com as normas técnicas para a elaboração do documento. Estes três professores são grandes exemplos desta sublime profissão.

Agradeço aos meus amigos da POLI pela ajuda fornecida durante o curso, em especial a Alessandra, Felipe, Gearlles, Nestor e Vinícius.

Por fim, agradeço a todos que contribuíram direta ou indiretamente na minha formação acadêmica.

Resumo

A LIPNet é uma rede neural piramidal utilizada com eficiência em visão computacional para reconhecer padrões. O primeiro conjunto de camadas dessa rede é responsável pela extração implícita de características; o segundo e último conjunto, por sua vez, realiza a classificação do padrão fornecido na entrada. Essa rede incorpora em sua arquitetura algumas motivações biológicas como o aprendizado profundo realizado no cérebro e a existência de campos receptivos e inibitórios do sistema visual humano. Adicionalmente, há outra teoria baseada no funcionamento das redes neurais do ser humano: o aprendizado por partes. Esse tipo de aprendizado busca entender um padrão a partir de conceitos mais simples que o formam. Estudos mostram que esse aprendizado pode ser alcançado introduzindo-se uma restrição não-negativa nos pesos da LIPNet. Dessa forma, a rede não-negativa conseguiria obter um aprendizado interpretável e, não, opaco, como acontece nas redes tradicionais. Assim, para aplicar essa restrição na rede, foi utilizada a Otimização por Enxame de Partículas (PSO). Além disso, foi criado um modelo de visualização do aprendizado da rede. Através deste foi possível comparar a representação interna do aprendizado do modelo não-negativo com a do modelo tradicional da LIPNet. Os resultados mostram que a rede neural com restrição não-negativa apresentou uma melhor interpretabilidade dos padrões fornecidos, embora tenha apresentado taxas de classificação ligeiramente inferiores.

Palavras-chave: LIPNet, Receptivo, Inibitório, PSO, Não-negatividade, Não-negativo, Aprendizado por partes.

Abstract

LIPNet is a pyramidal neural network efficiently used in computer vision for pattern recognition. The first layer set of the network is responsible for the implicit feature extraction; the second and final layer set, then, classifies the input patterns. This network incorporates in its architecture some biological motivations such as the deep-learning held in the brain and the existence of receptive and inhibitory fields in the human visual system. Additionally, there is another theory based on the functioning of human neural networks: parts-based learning. This type of learning aims to understand a pattern starting with its simplest concepts. Studies have shown that this learning can be achieved by introducing a non-negative constraint on the LIPNet weight. This way, the non-negative network could get an interpretable learning and, not, opaque, as it happens with traditional networks. Thus, in order to apply this constraint on the network, the Particle Swarm Optimization (PSO) algorithm was used. A display model of the network learning was also created. Through this model it was possible to compare the internal representation of the non-negative model learning and the traditional LIPNet model. The results show that the neural network with non-negative constraint presented a better interpretability of the patterns provided, although it achieved lower classification rates.

Keywords: LIPNet, Receptive, Inhibitory, PSO, Non-negativity, Non-negative, Parts-based learning.

Lista de ilustrações

Figura 1 – Imagens formadas pelos pesos da MLP sem restrição, à esquerda, e da MLP com restrição não-negativa, à direita.	16
Figura 2 – Relação entre os parâmetros da LIPNet: r é o tamanho do campo receptivo e o é <i>overlap</i> entre os campos receptivos adjacentes. O tamanho da região não sobreposta entre dois campos receptivos adjacentes, g , é calculado através da relação $g = r - o$	18
Figura 3 – Pseudocódigo do PSO.	20
Figura 4 – Funcionamento da LIPNet + PSO.	23
Figura 5 – Pseudocódigo do funcionamento da LIPNet + PSO.	23
Figura 6 – Representação do modelo de visualização do aprendizado LIPNet.	25
Figura 7 – Pseudocódigo do funcionamento do método principal do modelo de visualização do aprendizado da LIPNet.	26
Figura 8 – (a) Imagem original de uma face da base AR. (b), (c) e (d) Essa imagem após a aplicação da correção γ , considerando $c=1$ e γ sendo 0,6, 0,4 e 0,3, respectivamente.	26
Figura 9 – Pseudocódigo do funcionamento do método <i>calculaSaídaPorPixel</i>	27
Figura 10 – Imagens de faces e não-faces da base CBCL podem ser vistas em (a) e (b), respectivamente. Assim como algumas imagens de faces de mulheres e homens da base AR podem ser vistas em (c) e (d), respectivamente.	29
Figura 11 – (a) e (c) correspondem às representações de 300 imagens aleatórias do banco de teste da CBCL geradas pelo modelo de visualização do aprendizado da rede para a LIPNet tradicional de não-face e face, respectivamente; (b) e (d) correspondem às representações de 300 imagens aleatórias do banco de teste da CBCL geradas pelo modelo de visualização do aprendizado da rede para a LIPNet não-negativa de não-face e face, respectivamente. A configuração utilizada nestes resultados foi a (4,1,3,0) , o fator = 0,1, 2500 iterações do PSO e 500 iterações da LIPNet.	31

Figura 12 – (a), (b) e (c) são representações de 300 imagens aleatórias de faces do banco de teste da CBCL geradas pelo modelo de visualização do aprendizado da rede para a LIPNet com restrição correspondentes aos fatores 0,1, 0,2 e 0,5, respectivamente; (d), (e) e (f) são representações de 300 imagens aleatórias de não-faces do banco de teste da CBCL geradas pelo modelo de visualização do aprendizado da rede para a LIPNet com restrição correspondentes aos fatores 0,1, 0,2 e 0,5, respectivamente. A configuração utilizada nestes resultados foi a (4,1,3,0) e 2500 iterações do PSO foram consideradas.	31
Figura 13 – (a), (b), (c), (d) e (e) são imagens que correspondem a representação de 300 imagens aleatórias de face do banco de teste da CBCL geradas pelo modelo de visualização do aprendizado da rede para a LIPNet com restrição correspondentes aos valores de iterações do PSO 10, 100, 500, 1500 e 2500, respectivamente. A configuração utilizada nestes resultados foi a (9,4,3,2) e o fator=0,1.	32
Figura 14 – (a), (b), (c), (d) e (e) correspondem à representação de 300 imagens aleatórias de face do banco de teste da CBCL geradas pelo modelo de visualização do aprendizado da rede para a LIPNet sem restrição correspondentes aos valores de iterações da rede 10, 50, 100, 250 e 500, respectivamente. A configuração utilizada nestes resultados foi a (4,1,3,0) e o fator=0,1.	32
Figura 15 – (a) e (b) correspondem à representação de 300 imagens aleatórias de faces da base de teste da CBCL considerando a configuração (4,1,3,0) geradas pelo modelo de visualização do aprendizado da rede para a LIPNet com restrição; (c) e (d) correspondem à representação de 300 imagens aleatórias de faces da base de teste da CBCL considerando a configuração (9,4,3,2) geradas pelo modelo de visualização do aprendizado da rede para a LIPNet com restrição. O fator utilizado para estas imagens foi 0,1 e 2500 iterações do PSO foram consideradas.	33
Figura 16 – (a) e (e) mostram as 10 imagens do banco de teste da AR de cada gênero que foram consideradas; (b) e (f) são imagens que correspondem à representação dessas imagens geradas pelo modelo de visualização do aprendizado da rede para a LIPNet tradicional; (c) e (g) são imagens que correspondem à representação dessas imagens geradas pelo modelo de visualização do aprendizado da rede para a LIPNet não-negativa; por fim, (d) e (h) são as médias das 10 imagens. A configuração utilizada nestes resultados foi a (14,1,3,2), o fator = 0,1, 2500 iterações do PSO e 500 iterações da LIPNet.	34

Figura 17 – (a) e (e) mostram as 5 imagens do banco de teste da AR de cada gênero que foram consideradas; (b) e (f) são imagens que correspondem à representação dessas imagens geradas pelo modelo de visualização do aprendizado da rede para a LIPNet tradicional; (c) e (g) são imagens que correspondem à representação dessas imagens geradas pelo modelo de visualização do aprendizado da rede para a LIPNet não-negativa; por fim, (d) e (h) são as médias das 5 imagens. A configuração utilizada nestes resultados foi a (14,1,3,2), o fator = 0,1, 2500 iterações do PSO e 500 iterações da LIPNet.	35
Figura 18 – (a), (b) e (c) são imagens correspondente a 10 mulheres do banco de teste da AR geradas pelo modelo de visualização do aprendizado da rede para a LIPNet com restrição correspondentes aos fatores 0,1, 0,2 e 0,5, respectivamente. A configuração utilizada nestes resultados foi a (14,1,3,2) e 2500 iterações do PSO foram consideradas.	36
Figura 19 – (a), (b), (c), (d) e (e) são imagens que correspondem a 10 homens do banco de teste da AR geradas pelo modelo de visualização do aprendizado da rede para a LIPNet com restrição correspondentes aos valores de iterações do PSO 10, 100, 500, 1500 e 2500, respectivamente. A configuração utilizada nestes resultados foi a (14,1,3,2) e o fator=0,1.	36
Figura 20 – (a), (b), (c), (d) e (e) são imagens que correspondem a 10 homens do banco de teste da AR geradas pelo modelo de visualização do aprendizado da rede para a LIPNet sem restrição correspondentes aos valores de iterações da rede 10, 50, 100, 250 e 500, respectivamente. A configuração utilizada nestes resultados foi a (14,1,3,2) e o fator=0,1.	37
Figura 21 – (a) e (b) correspondem à representação de 10 imagens de mulheres e homens, respectivamente, da base de teste da AR considerando a configuração (4,1,3,0) geradas pelo modelo de visualização do aprendizado da rede para a LIPNet com restrição; (c) e (d) correspondem à representação de 10 imagens de mulheres e homens, respectivamente, da base de teste da AR considerando a configuração (14,1,3,2) geradas pelo modelo de visualização do aprendizado da rede para a LIPNet com restrição. O fator utilizado para estas imagens foi 0,1 e 2500 iterações do PSO foram consideradas.	37

Lista de tabelas

Tabela 1 – Variáveis do PSO.	20
Tabela 2 – Componentes da partícula do PSO do modelo não-negativo.	22
Tabela 3 – Arranjos experimentais.	30
Tabela 4 – Valores da AUC encontrados nos experimentos realizados com a CBCL.	30
Tabela 5 – Valores da AUC encontrados nos experimentos realizados com a AR.	33

Lista de abreviaturas e siglas

AUC	<i>Area Under the Curve</i>
CBCL	<i>Center for Biological and Computational Learning</i>
CE	<i>Cross Entropy</i>
LIPNet	<i>Lateral Inhibition Pyramidal Neural Network</i>
MLP	<i>Multilayer Perceptron</i>
NMF	<i>Non-negative Matrix Factorization</i>
PC	<i>Principal Components</i>
PCA	<i>Principal Components Analysis</i>
PSO	<i>Particle Swarm Optimization</i>
RNA	<i>Rede Neural Artificial</i>
ROC	<i>Receiver Operating Characteristic</i>
Rprop	<i>Resilient Propagation</i>
VQ	<i>Vector Quantization</i>

Sumário

1	INTRODUÇÃO	13
1.1	Qualificação do Problema	13
1.2	Objetivos	14
1.2.1	Objetivos Específicos	14
1.3	Estrutura da Monografia	14
2	REFERENCIAL TEÓRICO	15
2.1	Introdução	15
2.2	Não-negatividade	15
2.3	LIPNet	16
2.3.1	Inspiração biológica	17
2.3.2	Arquitetura e treinamento da LIPNet	18
2.4	PSO	19
3	LIPNET NÃO-NEGATIVA	22
3.1	Funcionamento da LIPNet não-negativa	22
3.2	Modelo de visualização do aprendizado da LIPNet	24
4	EXPERIMENTOS E RESULTADOS	28
4.1	Metodologia	28
4.2	CBCL	30
4.3	AR	33
4.4	Principais considerações	36
5	CONSIDERAÇÕES FINAIS	39
5.1	Conclusões	39
5.2	Trabalhos Futuros	39
	REFERÊNCIAS	41

1 Introdução

Neste trabalho de conclusão de curso, foi proposto um modelo de LIPNet (do inglês, *Lateral Inhibition Pyramidal Neural Network*) que realiza um aprendizado por partes, através da incorporação da não-negatividade na rede, para o reconhecimento de padrões visuais. Para este fim, foi utilizada inteligência de enxames para apenas permitir pesos positivos na rede. Neste capítulo, organizado em quatro seções, apresenta-se a introdução do documento. A Seção 1.1 apresenta a motivação para a realização deste trabalho. Já na Seção 1.2 são descritos os objetivos gerais e específicos para a solução do problema. Por fim, na Seção 1.3 é relatada a estrutura do restante do documento.

1.1 Qualificação do Problema

Visão computacional é uma área da computação que busca realizar, através de máquinas, a capacidade dos seres humanos de interpretar uma imagem ou conjunto de imagens. Como aplicações importantes desse domínio cita-se o reconhecimento de escrita (GOUVEIA et al., 2014; BEZERRA; ZANCHETTIN; ANDRADE, 2012), de faces (FERNANDES; CAVALCANTI; REN, 2011) e a detecção de faces (SHARIFARA; RAHIM; ANISI, 2014). Aspectos provenientes do sistema visual humano são muitas vezes introduzidos em modelos dessa área a fim de que haja o reconhecimento de padrões nas imagens a serem interpretadas.

Neste contexto de inserção de inspirações biológicas, cita-se a Rede Neural Artificial (RNA), técnica da área de visão computacional que busca reproduzir em pequena escala o funcionamento do cérebro humano. Este modelo tem sido aplicado com êxito na tarefa de identificar padrões visuais (PEREZ et al., 2003; LECUN et al., 1998; NOLKER; RITTER, 2002).

Baseados na teoria da divisão de camadas hierárquicas com diferentes responsabilidades no cérebro, surgiram os modelos de arquitetura profunda (BENGIO et al., 2007). RNAs desse tipo apresentam vários níveis de abstração e permitem que o aprendizado da rede comece a partir de conceitos mais simples e evolua até a classificação dos padrões.

Outra inspiração proveniente do cérebro humano utilizada em RNAs é a existência de campos receptivos e inibitórios. O efeito gerado pela inserção de campos receptivos na arquitetura da rede é a extração implícita de características. Por sua vez, os campos inibitórios trazem maior estabilidade e eficácia para as RNAs (FERNANDES, 2013).

Adicionalmente, surge o aprendizado por partes como motivação para criação de novos tipos de RNAs. Este artifício é utilizado pelo cérebro humano para entender

relações complexas. Quando um novo padrão é apresentado, o cérebro inicialmente busca entender os conceitos mais simples que o formam para então interpreta-lo completamente. Estudos mostram que RNAs que incorporam restrições não-negativas em seu funcionamento realizam um aprendizado por partes (CHOROWSKI; ZURADA, 2015), o qual também gera melhorias na interpretabilidade da rede.

LIPNet é um tipo de RNA piramidal que utiliza em sua arquitetura os conceitos de aprendizagem profunda e campos receptivos e inibitórios (FERNANDES, 2013). O aprendizado por partes obtido através da incorporação da não-negatividade pode fazer com que esta rede obtenha uma melhor interpretação dos padrões apresentados a ela.

1.2 Objetivos

O objetivo geral deste trabalho é analisar o efeito do aprendizado por partes através da restrição não-negativa de pesos na LIPNet para o reconhecimento de padrões visuais. Para isso, um modelo de visualização do aprendizado da rede será implementado e utilizado para realizar as comparações entre o modelo tradicional e o não-negativo da LIPNet. Este último será elaborado com o auxílio da técnica de otimização *Particle Swarm Optimization* (PSO) (KENNEDY; EBERHART, 1995; ENGELBRECHT, 2007), visto que é apropriada para otimizar variáveis contínuas em espaços hiperdimensionais.

1.2.1 Objetivos Específicos

Por objetivos específicos, cita-se:

- Estudo do estado-da-arte da não-negatividade, da LIPNet e do PSO;
- Implementação de modelo para visualização do aprendizado da rede;
- Realização de experimentos em duas bases de imagens de faces;
- Análise e comparação dos resultados obtidos.

1.3 Estrutura da Monografia

Este trabalho está estruturado em 5 capítulos. O Capítulo 2 apresenta a fundamentação teórica ressaltando a não-negatividade, LIPNet e o PSO. No Capítulo 3 o funcionamento da LIPNet com pesos não-negativos é detalhado, assim como o modelo de visualização do aprendizado da rede. Em seguida, o Capítulo 4 apresenta os experimentos e os resultados obtidos. Finalmente, o Capítulo 5 expõe as conclusões e propõe alguns trabalhos futuros.

2 Referencial Teórico

2.1 Introdução

Neste capítulo são abordados os três principais conceitos que envolvem este trabalho. Na Seção 2.2, a não-negatividade é apresentada assim como modelos que aplicam esta abordagem. Em seguida, na Seção 2.3 a LIPNet é detalhada. Por fim, a Seção 2.4 expõe de forma resumida o funcionamento do PSO.

2.2 Não-negatividade

Não-negatividade é uma condição natural para várias aplicações do mundo real, como economia, bioinformática e visão computacional (ZASS; SHASHUA, 2007). Devido a características físicas inerentes aos sistemas sob investigação, a não-negatividade é usualmente imposta nos parâmetros a fim de prevenir absurdos físicos e resultados não-interpretáveis (CHEN et al., 2011; VU et al., 2015). Representações não-negativas também estão relacionadas à não-negatividade das taxas de disparo neurais da modelagem biológica (CHEN et al., 2014; XIUJUN; CHANG, 2014).

Baseada em diferentes motivações, a não-negatividade têm sido incorporada a técnicas como RNAs (FYFE, 1997; XU, 1993; YASUDA, 2014). Em razão da existência da teoria cerebral de que os sinais dos neurônios podem ser excitatórios ou inibitórios, mas não ambos, Fyfe (FYFE, 1997) propôs um modelo de rede com pesos não-negativos para encontrar os componentes principais (PC, do inglês *Principal Components*). Lemme et al (LEMME; REINHART; STEIL, 2010) aplicou um *autoencoder* modificado com restrições não-negativas para produzir uma codificação não-negativa de imagens de entrada. Estudos demonstram que pesos não-negativos produzem estabilidade em redes assimétricas de Hopfield (MA, 1997), propriedade anteriormente encontrada apenas em redes recorrentes simétricas. Além disso, RNAs com pesos positivos estão sendo desenvolvidas para obter modelos monótonos (DANIELS; VELIKOVA, 2010; DANIELS; KAMP, 1999), utilizados para resolver problemas nos quais é necessário que o valor da saída cresça à medida que o da entrada cresce.

Pesquisas também apontam a não-negatividade como um dos fatores responsáveis pelo aprendizado por partes realizado por RNAs (CHOROWSKI; ZURADA, 2015; GE; IWATA, 2002). Ao utilizar uma base de imagens de faces nos experimentos, Xijin e Shui-chi (GE; IWATA, 2002) mostram que a rede autoassociativa não-negativa implementada por eles descobriu características localizadas que se assemelham com noções intuitivas

de partes de faces. O modelo desenvolvido por eles, utiliza apenas a noção de campos receptivos, diferentemente do funcionamento do cérebro humano, no qual existe a presença de efeitos inibitórios. Chorowski e Zurada ([CHOROWSKI; ZURADA, 2015](#)) demonstram que restringir o intervalo dos pesos para não-negativo melhora a interpretabilidade da rede *Multilayer Perceptron* (MLP). O experimento com a base de dígitos manuscritos MNIST mostra que uma das consequências do uso dessa restrição é a geração de pesos mais esparsos na camada de classificação. Adicionalmente, os padrões aprendidos pelos neurônios ocultos são localizados e tendem a ser mais similares a partes de dígitos, como pode ser visto na Figura 1.

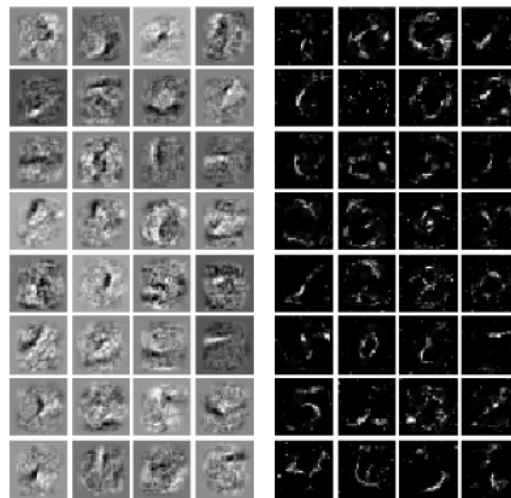


Figura 1 – Imagens formadas pelos pesos da MLP sem restrição, à esquerda, e da MLP com restrição não-negativa, à direita.

[Fonte: reproduzido de ([CHOROWSKI; ZURADA, 2015](#))]

A técnica conhecida como *Nonnegative matrix factorization* (NMF) realiza um aprendizado por partes empregando restrições não-negativas [13], [14]. Esta restrição é a principal diferença entre este método e outros como o *Principal Component Analysis* (PCA) e o *Vector Quantization* (VQ), já que o NMF só produz vetores base não-negativos. A restrição não-negativa imposta neste modelo é consistente com a noção de combinar partes para formar o todo, já que esta permite apenas combinações aditivas. A não-negatividade também tem sido empregada em outros métodos de redução de dimensionalidade ([ZHANG; YU, 2013](#); [PLUMBLEY; OJA, 2004](#)), grafos ([ZHUANG et al., 2012](#); [HE et al., 2011](#)) e reconstrução de dados ([LUAN et al., 2014](#); [ZHANG; LI; HUANG, 2013](#)).

2.3 LIPNet

O objetivo das RNAs é reproduzir, mesmo que em pequena escala, o funcionamento do cérebro humano. Dessa forma, é comum que autores busquem inspirações biológicas

para criar novos tipos de redes. A LIPNet é uma rede piramidal que incorpora o conceito de campos receptivos e inibitórios baseados no sistema visual humano. Baseada na PyraNet (PHUNG; BOUZERDOUM, 2007a), esta RNA foi criada por (FERNANDES, 2013) e será detalhada nesta seção. Na Subseção 2.3.1 é apresentada a inspiração biológica que motivou sua criação. Em seguida, a Subseção 2.3.2 descreve sua arquitetura e seu funcionamento.

2.3.1 Inspiração biológica

Redes de arquitetura profunda são baseadas na teoria de aprendizado profundo realizado pelo cérebro. Este aprendizado sugere a existência de camadas hierárquicas com diferentes responsabilidades no cérebro. RNAs desse tipo necessitam de menos padrões para treinamento visto que seu aprendizado ocorre através de múltiplas camadas de processamento. Esses modelos conseguem generalizar o aprendizado para padrões não vistos anteriormente. Entre os modelos de arquitetura profunda aplicados na área de visão computacional destacam-se o *Neocognitron* (FUKUSHIMA; MIYAKE; ITO, 1983), a rede neural convolucional (LECUN et al., 1989) e a rede neural piramidal (PHUNG; BOUZERDOUM, 2007b).

Campo receptivo pode ser definido como uma área que, ao receber um estímulo apropriado, ativa a resposta de um determinado neurônio (LEVINE; SHEFNER, 1991). Esta área delimita um grupo de neurônios que se conecta a outro neurônio localizado hierarquicamente acima do grupo. Esse neurônio posicionado acima é responsável por extrair características, tais como bordas, do grupo ao qual ele está ligado. Dessa forma, os campos receptivos são utilizados em RNAs para realizar a extração implícita de características dos dados brutos fornecidos na entrada. Algumas modelos neurais que aplicam a noção de campos receptivos em suas arquiteturas podem ser vistos em (PARK; PEDRYCZ; OH, 2009; GHOSH; PAL, 2010).

A presença de outra área que atua de forma inibitória, simultaneamente ao campo receptivo, sobre a saída do neurônio deu origem ao conceito de campo inibitório (RIZZOLATTI; CAMARDA, 1975). A inserção de campos inibitórios torna as RNAs mais estáveis e eficazes (MAO; MASSAQUOI, 2007), além de gerar ganhos na análise de textura (GRIGORESCU; PETKOV; WESTENBERG, 2003).

Neste cenário, surge a LIPNet, uma rede neural piramidal que aplica os conceitos de arquitetura profunda e campos receptivos e inibitórios. Esta RNA pode ser considerada como uma reformulação da PyraNet que implementa o conceito de inibição lateral. Este modelo vem apresentando resultados satisfatórios em problemas de visão computacional (SOARES; FERNANDES; BASTOS-FILHO, 2014; FERNANDES; CAVALCANTI; REN, 2013).

2.3.2 Arquitetura e treinamento da LIPNet

A LIPNet possui uma estrutura formada por duas camadas, assim como a PyraNet. A primeira camada tem uma organização similar a uma pirâmide, com uma estrutura 2D, na qual cada neurônio está conectado a um campo receptivo com neurônios da camada anterior. Nesta, é realizada a extração implícita de características do padrão fornecido e também a redução da dimensionalidade. Na segunda camada, com uma estrutura 1D, faz-se a classificação das imagens baseada nas características encontradas na camada anterior. Esta camada 1D funciona de forma similar a uma MLP totalmente conectada.

O relacionamento dos neurônios da camada 2D ocorre da seguinte maneira: um neurônio de uma determinada camada se liga aos neurônios que estão em seu campo receptivo na camada anterior. *Overlap* é o termo dado quando há sobreposição entre campos receptivos adjacentes e pode ser visualizado através da Figura 2. A inibição lateral ocorre quando existe uma interferência negativa proveniente de um neurônio de origem na saída de um neurônio de destino. Para isso acontecer, os neurônios precisam estar dentro de um mesmo campo receptivo e na mesma camada.

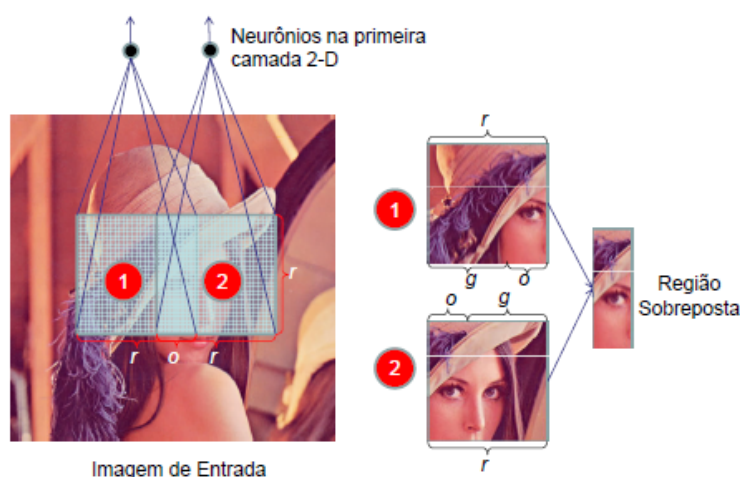


Figura 2 – Relação entre os parâmetros da LIPNet: r é o tamanho do campo receptivo e o é *overlap* entre os campos receptivos adjacentes. O tamanho da região não sobreposta entre dois campos receptivos adjacentes, g , é calculado através da relação $g = r - o$.

[Fonte: reproduzido de (FERNANDES, 2013)]

Na LIPNet, o peso de um neurônio da camada 1D é associado entre um neurônio de uma camada e outro da camada posterior. Em contrapartida, o peso de um neurônio da camada 2D é associado com a camada posterior, ou seja, pesos associados aos próprios neurônios. Este funcionamento faz com que a quantidade de memória necessária para memorizar a rede e o seu custo computacional reduzam (FERNANDES, 2013).

A saída de um neurônio 2D é dada pela aplicação de uma função de ativação

não-linear sobre o somatório ponderado dos neurônios dentro do seu campo receptivo subtraído pelo somatório ponderado dos neurônios que estão em seu campo inibitório. Da mesma forma, a saída de um neurônio 1D consiste na aplicação de uma função de ativação não-linear sobre o somatório ponderado pelos pesos das conexões dos neurônios na camada anterior.

A primeira camada 2D possui o tamanho da imagem fornecida na entrada, visto que cada neurônio desta representa um pixel. Já a saída da rede tem como quantidade de neurônios o número de classes do problema em questão e leva em consideração os estímulos excitatório e inibitórios dos neurônios vizinhos.

O método aplicado para ajustar os pesos é o *Resilient Propagation* (Rprop) (RIEDMILLER; BRAUN, 1993), bastante utilizado na literatura devido aos bons resultados fornecidos. Além disso, como métrica do sucesso da LIPNet é usada a área sob a curva ROC (AUC - do inglês, *Area Under the Curve*) (BRADLEY, 1997). Esta mede a precisão obtida fornecendo um valor dentro do intervalo $[0,1]$, no qual 1 é o melhor resultado.

2.4 PSO

PSO é um algoritmo de otimização por enxame de partículas que tenta simular o comportamento social de grupo de pássaros. A técnica, pertencente à área de Computação Natural, é modelada através de partículas que fazem uso de sua experiência e da do próprio bando para encontrar a melhor região do espaço de busca. Desde sua publicação inicial em 1995 (KENNEDY; EBERHART, 1995), outras variações do PSO já foram propostas com o objetivo de refinar a técnica ou modificá-la para resolver problemas específicos (ABDULLAH et al., 2012; FENG; PAN, 2014; SONG et al., 2014).

O PSO possui uma população de partículas, onde cada partícula representa uma possível solução para o problema de otimização. Cada partícula do exame pode ser representada por um objeto que possui associado a ele um vetor posição, isto é, a localização dela no espaço de busca, e um vetor velocidade, responsável por guiar as mudanças da posição das partículas. O objetivo é fazer com que as partículas se movimentem pelo espaço de busca, ou seja, mudem suas posições, tendo suas velocidades atualizadas dinamicamente de acordo com o histórico das experiências individuais e coletivas vividas. Assim, o algoritmo evolui à medida que o enxame tenta percorrer todo o espaço de busca para encontrar a melhor solução para o problema.

Em geral, o enxame é inicializado com os valores dos vetores de velocidade e posição gerados aleatoriamente. A primeira iteração do algoritmo inicia com a atribuição de valores aos parâmetros da equação de velocidade. Definem-se então os valores referentes ao enxame, constantes e o critério de parada. Para encontrar a solução, a técnica utiliza uma função de avaliação, conhecida como *fitness*, para guiar as partículas durante a busca. Esta é

específica para cada problema e pode ser uma função de máximo ou mínimo. Considerando as variáveis exibidas na Tabela 1, o funcionamento do PSO básico pode ser descrito através do pseudocódigo exibido na Figura 3.

Tabela 1 – Variáveis do PSO.

[Fonte: elaboração própria]

Variáveis	Descrição
\vec{x}_i	Posição no espaço de busca
\vec{v}_i	Velocidade no espaço de busca
\vec{p}_i	Melhor posição da partícula i no instante atual
\vec{n}_i	Melhor posição da vizinhança da partícula i no instante atual

```

Inicializar população
Até atingir um critério de parada
  Para cada partícula  $i$  do enxame
    Atualizar velocidade  $\vec{v}_i$ 
    Atualizar posição  $\vec{x}_i$ 
    Avaliar  $fitness$ 
    Se melhor posição individual melhorou
      Atualiza  $\vec{p}_i$ 
    Fim
    Se melhor posição da vizinhança  $\vec{n}_i$ , melhorou
      Atualiza  $\vec{n}_i$ 
    Fim
  Fim
Fim

```

Figura 3 – Pseudocódigo do PSO.

[Fonte: elaboração própria]

A atualização da posição da partícula é dada pela seguinte equação:

$$\vec{x}_i(t+1) = \vec{x}_i(t) + \vec{v}_i(t+1). \quad (2.1)$$

Assim como a atualização da velocidade pode ser obtida através da fórmula:

$$\vec{v}_i(t+1) = w\vec{v}_i(t) + c_1r_1[\vec{p}_i(t) - \vec{x}_i(t)] + c_2r_2[\vec{n}_i(t) - \vec{x}_i(t)]. \quad (2.2)$$

O parâmetro w é chamado de fator de inércia, um número entre 0 e 1, que tem como objetivo controlar a velocidade para que uma partícula muito rápida não passe direto por uma solução ideal. O valor de w pode ser regulado, dependendo do objetivo, de forma que não fique muito próximo de 1, para que as velocidades não tenham valores altos, nem fique muito próximo de 0, para que as partículas fiquem sem muito movimento, não

realizando uma busca completa pelo espaço. As variáveis r_1 e r_2 são valores aleatórios dentro do intervalo $[0, 1]$ e c_1 e c_2 são duas constantes positivas que definem a influência do conhecimento da partícula e o do enxame, respectivamente, no cálculo da velocidade.

Há ainda uma variação no cálculo da velocidade, proposta em 2002 (CLERC; KENNEDY, 2002), que introduz o fator de constrição de Clerc (X). A fórmula da velocidade seria então alterada para:

$$\vec{v}_i(t+1) = X[\vec{v}_i(t) + c_1 r_1 (\vec{p}_i(t) - \vec{x}_i(t)) + c_2 r_2 (\vec{n}_i(t) - \vec{x}_i(t))]. \quad (2.3)$$

O valor mais utilizado na literatura para este fator é 0,729.

Assim como os valores utilizados nos parâmetros do PSO influenciam o resultado do algoritmo, a escolha da topologia também gera diferentes respostas. A topologia do PSO define a forma como ocorrerá a comunicação entre as partículas. Existem vários tipos de topologia, porém as mais usadas na literatura são a global e a local. No modelo global, todas as partículas estão conectadas entre si. Assim, todas elas possuem o conhecimento de qual é a melhor posição já encontrada pelo enxame, possibilitando uma convergência mais rápida. Entretanto, esta topologia pode provocar uma convergência prematura. A local em formato anel permite apenas dois vizinhos a cada partícula. Embora a troca de informação entre as partículas seja mais lenta, esta estrutura provê uma busca mais completa. Esta última será a topologia considerada neste trabalho.

3 LIPNet não-negativa

Este capítulo aborda o modelo proposto a fim de obter um aprendizado por partes na LIPNet. Para isso, foi implementada uma restrição não-negativa nos pesos da RNA utilizando o PSO. Assim, o funcionamento da LIPNet não-negativa será descrito na Seção 3.1.

Com o objetivo de comparar o aprendizado realizado pela LIPNet tradicional e a não-negativa, foi implementado um modelo de visualização do aprendizado. Este é apresentado na Seção 3.2.

3.1 Funcionamento da LIPNet não-negativa

O PSO é um algoritmo que pode ser utilizado para otimizar redes neurais e já se mostrou eficaz quando aplicado na LIPNet (SOARES; FERNANDES; BASTOS-FILHO, 2014). No caso deste trabalho, o PSO foi usado com o objetivo principal de implementar a restrição não-negativa nos pesos da RNA. Assim, o método tradicional de ajuste dos pesos da LIPNet, Rprop, foi substituído, por sua vez, pelos pesos encontrados pelo método de otimização. A partícula deste modelo é composta pelos pesos e pelas polarizações externas essenciais para a LIPNet. Na Tabela 2 é possível visualizar os componentes de uma partícula do enxame.

Tabela 2 – Componentes da partícula do PSO do modelo não-negativo.

[Fonte: elaboração própria]

Componente	Descrição
$w_{i,j}^l$	Pesos da camada 2D, onde $i \in \{1, \dots, H^l\}$, $j \in \{1, \dots, W^l\}$ e $l \in \{1, \dots, L_{2D}\}$
$b_{i,j}^l$	Polarizações externas da camada 2D, onde $i \in \{1, \dots, H^l\}$, $j \in \{1, \dots, W^l\}$ e $l \in \{1, \dots, L_{2D}\}$
$w_{m,n}^l$	Pesos da camada 1D, onde $m \in \{0, \dots, L_{1D} - 1\}$, $n \in \{1, \dots, L_{1D}\}$ e $l \in \{1, \dots, L_{1D}\}$
b_m^l	Polarizações externas da camada 1D, onde $m \in \{0, \dots, L_{1D} - 1\}$, $n \in \{1, \dots, L_{1D}\}$ e $l \in \{1, \dots, L_{1D}\}$

Inicialmente, uma configuração da LIPNet é definida como entrada para o PSO. O objetivo é encontrar a solução ótima que determinará o conjunto de pesos a serem utilizados pela rede que geram a máxima taxa de classificação, medida através da AUC. Assim, o treinamento da rede é a própria otimização. A cada iteração, o *fitness* é calculado,

os pesos são reajustados e os melhores valores guardados para serem retornados ao fim do processo de otimização. A relação entre o PSO e a LIPNet pode ser verificada na Figura 4.

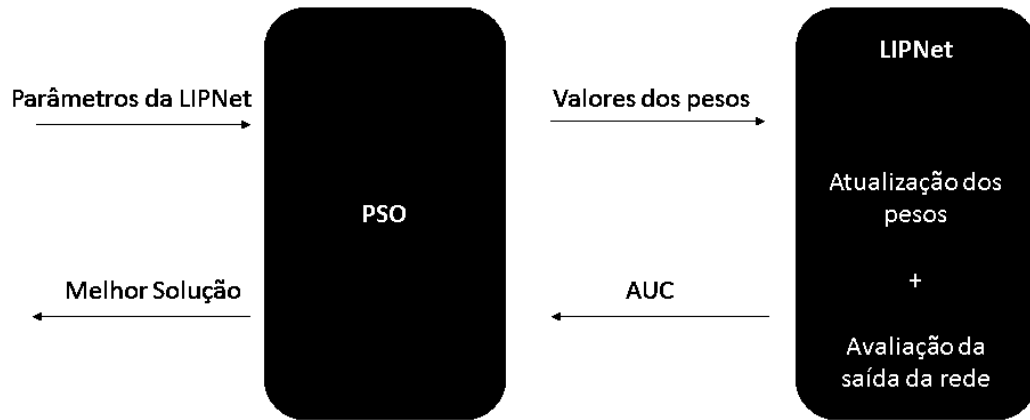


Figura 4 – Funcionamento da LIPNet + PSO.

[Fonte: adaptado de (SOARES; FERNANDES; BASTOS-FILHO, 2014)]

O funcionamento da otimização pode ser visualizado através do pseudocódigo exibido na Figura 5. Nesta, \vec{n}_{best} corresponde à melhor solução encontrada pelo enxame e as outras variáveis foram descritas na Seção 2.4 do capítulo anterior.

```

Inicializar população
Até atingir um critério de parada
  Para cada partícula  $i$  do enxame
    Atualizar velocidade  $\vec{v}_i$ 
    Atualizar posição  $\vec{x}_i$ 
     $lipnet.atualizaPesos(\vec{x}_i)$ 
     $fitness_i = lipnet.avaluaRede()$ 
    Atualizar  $\vec{p}_i$ 
  Fim
  Para cada partícula  $i$  do enxame
    Atualiza  $\vec{n}_i$ 
  Fim
  Atualiza  $\vec{n}_{best}$ 
Fim
Retorna  $\vec{n}_{best}$ 

```

Figura 5 – Pseudocódigo do funcionamento da LIPNet + PSO.

[Fonte: elaboração própria]

No PSO, os intervalos dos parâmetros que compõem as partículas podem ser definidos. Dessa forma, é possível gerar valores de pesos apenas positivos. O intervalo utilizado para os pesos da rede foi $[10^{-8}, 100]$, o qual abrange a maioria dos pesos gerados pela LIPNet original e foi usado em (SOARES; FERNANDES; BASTOS-FILHO, 2014).

Já na LIPNet, como função de ativação, apenas a sigmóide-logística foi utilizada no treinamento, visto que esta gera valores no intervalo $[0,1]$, ou seja, não-negativos. Vale

ressaltar também que o valor do *bias* não foi restrito a valores não-negativos. Além disso, o método de entropia cruzada (CE - *Cross Entropy*) (BISHOP, 1995) é usado na saída da rede. Este calcula a probabilidade *a posteriori* para cada uma das classes existentes. Sendo assim, a imagem será classificada como pertencente à classe na qual ela possui a maior probabilidade de pertencer. Este método foi utilizado na LIPNet não-negativa pois permite a convergência mesmo que o sinal dos pesos esteja restrito a apenas positivo (CHOROWSKI; ZURADA, 2015).

3.2 Modelo de visualização do aprendizado da LIPNet

Para realizar um comparativo entre o aprendizado da LIPNet tradicional e desta com a restrição, foi criado um modelo de visualização do aprendizado. As imagens geradas por este modelo são formadas observando-se a influência que cada neurônio obteve para a classificação da rede, como mostra a Figura 6.

Cada neurônio da RNA influencia em certo grau na saída. Para identificar uma face, por exemplo, espera-se que alguns neurônios responsáveis pela região dos olhos sejam mais excitados. Ao formar uma imagem com o que cada neurônio aprendeu do padrão, visualiza-se o aprendizado geral da LIPNet, ou seja, que partes foram mais consideradas para a tomada de decisão em relação à classificação do mesmo. Assim, para uma imagem fornecida na entrada, calcula-se a saída de cada neurônio e monta-se uma imagem normalizada com esses valores. Dado um determinado neurônio, o cálculo leva em consideração os neurônios conectados a ele.

O pseudocódigo exibido na Figura 7 descreve o funcionamento do método que realiza a montagem das imagens. A função *signum* retorna o sinal do número passado, onde 0 é o retorno quando o número é 0, 1 é maior que 0 e -1 se é menor que 0. Os valores considerados para *fator* variaram e estão descritos na Seção 4. Essa potenciação foi acrescentada para obter uma melhor visualização na imagem gerada. O conceito base utilizado é a lei da potência (ou correção *gamma*) de realce no domínio do espaço, da área de Processamento Digital de Imagens (GONZALEZ; WOODS, 2011). Essa lei, na prática, é a aplicação de uma potenciação em cada pixel da imagem. Quanto maior o valor de *gamma* ou *fator*, como foi chamado neste trabalho, mais escura a imagem será. Um exemplo da utilização dessa lei pode ser visualizado na Figura 8.

Na Figura 7, o método *calculaSaídaPorPixel* implementa o cálculo da saída considerando os neurônios conectados, a cada camada, ao neurônio em questão. O pseudocódigo desse método é descrito na Figura 9. Primeiramente, cria-se uma matriz de variáveis *booleanas* formada pelos neurônios conectados ao neurônio em questão. Em seguida, atualiza-se a saída para os neurônios conectados nas camadas 2D ao neurônio considerado. Então, transforma-se as saídas dos neurônios 2D para serem entradas das camadas 1D. Após isso,

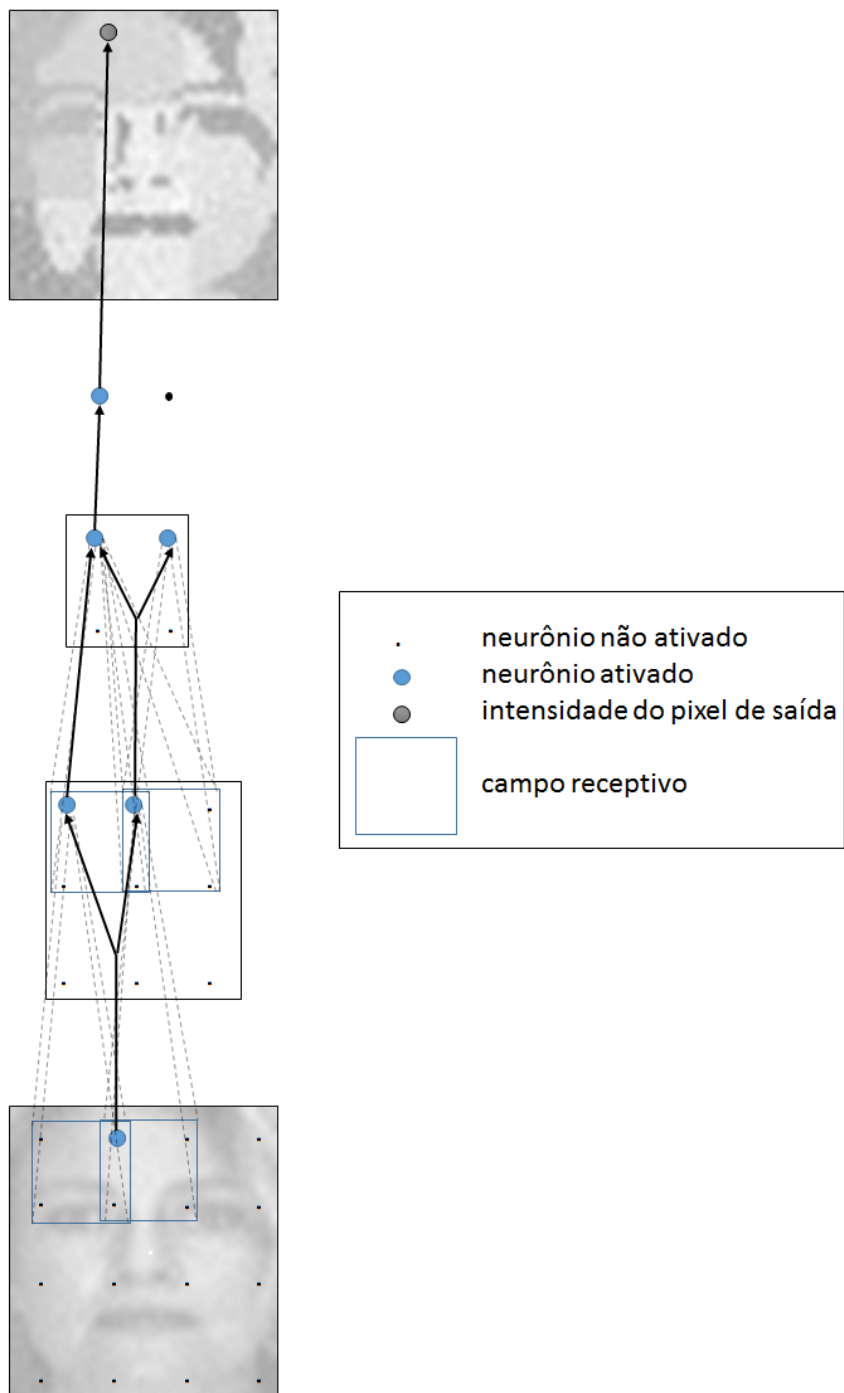


Figura 6 – Representação do modelo de visualização do aprendizado LIPNet.

[Fonte: elaboração própria]

```

Para cada imagem i
  Para cada pixel j
    saída = calculaSaídaPorPixel(imagem(i),pixel(j))
    x = signum(saída) * |saída|(fator)
    imagemGerada += x
  Fim
  normaliza(imagemGerada)
Fim
Retorna imagemGerada

```

Figura 7 – Pseudocódigo do funcionamento do método principal do modelo de visualização do aprendizado da LIPNet.

[Fonte: elaboração própria]

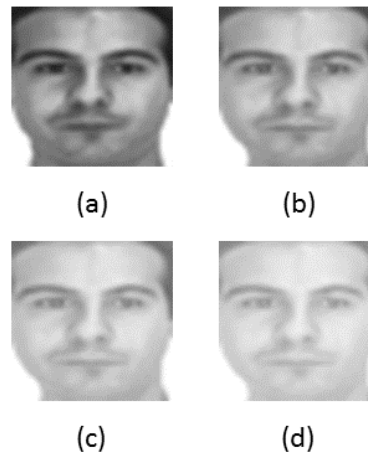


Figura 8 – (a) Imagem original de uma face da base AR. (b), (c) e (d) Essa imagem após a aplicação da correção *gamma*, considerando $c=1$ e γ sendo 0,6, 0,4 e 0,3, respectivamente.

[Fonte: elaboração própria.]

atualiza-se a saída para os neurônios conectados nas camadas 1D ao neurônio considerado. Vale ressaltar que tanto para os neurônios da camada 2D como para os da 1D, não é considerada a função de ativação nem o CE, apenas o somatório do produto do peso pela saída, visto que o objetivo é observar a influência de cada neurônio para o aprendizado dos padrões. Por fim, as saídas são passadas para um *array* de retorno.

```
neuroniosConsiderados = pixel;

Para cada camada2D k
    Se for a primeira camada2D
        neuroniosConsiderados = atualizaSaidaNeuronios (imagem, neuroniosConsiderados)
    Se não for
        neuroniosConsiderados = atualizaSaidaNeuronios (Camada(k-1).getNeuronios(), neuroniosConsiderados)
    Fim
Fim

entradas1D = rearrumaSaida()

Para cada neuronioConsiderado i
    neuroniosConsiderados1D[i] = transformaSaida2Dpara1D(neuronioConsiderado[i])
Fim

Para cada camada1D k
    camada1D[k].atualizaSaidaNeuronios(entradas1D, neuroniosConsiderados1D)
Fim

resp = transformaSaidaemArray()

Retorna resp
```

Figura 9 – Pseudocódigo do funcionamento do método *calculaSaidaPorPixel*.

[Fonte: elaboração própria]

4 Experimentos e Resultados

Este capítulo relata os resultados obtidos neste projeto. Na Seção 4.1 é detalhada a metodologia utilizada para realizar os experimentos. Já as Seções 4.2 e 4.3 apresentam os resultados e a análise referente às duas bases aplicadas nos testes. A Seção 4.4 traz algumas considerações sobre os resultados alcançados.

4.1 Metodologia

O aprendizado por partes em redes neurais ocorre quando esta busca aprender partes mais simples de um padrão para então interpretá-lo como um todo. Para analisar este efeito nesta pesquisa, foram utilizadas bases de imagens de faces, esperando-se que o modelo não-negativo buscasse focar nas partes mais importantes utilizadas para identificar uma face.

Para a realização do reconhecimento de faces a base de imagens criada pelo *Center for Biological and Computational Learning* (CBCL) (HEISELE; POGGIO; PONTIL., 2000) foi utilizada. Esta base possui um total de 6977 imagens de treino e 24045 de teste divididas em faces e não-faces, todas de tamanho 19x19 e em tons de cinza. Nesta pesquisa, utilizou-se a seguinte divisão das imagens: 1000 imagens de treino de faces e 1000 de não-faces, 1429 imagens para validação-cruzada de faces e 3544 de não-faces e 472 de teste de faces e 23573 de não faces. As imagens tiveram seus histogramas equalizados.

Adicionalmente a base AR Face (MARTINEZ; BENAVENTE, 1998) foi aplicada no reconhecimento de gênero. Desta base só foram utilizadas 1400 imagens, referentes às imagens sem oclusão, redimensionadas para 40x40 e em tons de cinza. Selecionou-se as primeiras imagens de cada pessoa para treino, totalizando 50 imagens de cada gênero, e as imagens número 14 de cada pessoa para teste, formando um total de 50 imagens de cada gênero também. Algumas imagens já pré-processadas das bases utilizadas podem ser vistas na Figura 10.

As imagens das bases de dados foram normalizadas utilizando-se o mesmo intervalo usado em (SOARES; FERNANDES; BASTOS-FILHO, 2014; FERNANDES, 2013). Testes foram realizados incluindo intervalos não-negativos, porém, os resultados obtidos não foram bons.

Como citado anteriormente, o algoritmo de inteligência de enxames PSO foi utilizado para implantar a restrição não-negativa na LIPNet. Utilizou-se a topologia local para a comunicação no enxame. Admitiu-se como valores do fator de constrição, coeficientes de aceleração cognitivo e social: 0,72984 e 2,05, respectivamente. A quantidade partículas

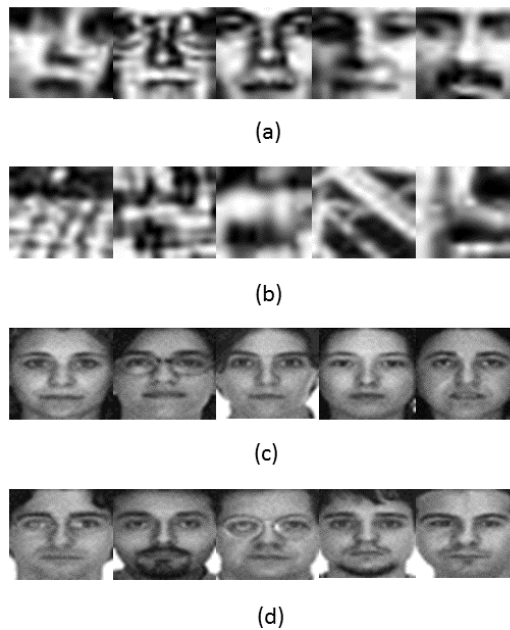


Figura 10 – Imagens de faces e não-faces da base CBCL podem ser vistas em (a) e (b), respectivamente. Assim como algumas imagens de faces de mulheres e homens da base AR podem ser vistas em (c) e (d), respectivamente.

[Fonte: elaboração própria]

consideradas no enxame foi 20, assim como em (SOARES; FERNANDES; BASTOS-FILHO, 2014). Como o critério de parada do PSO, assim como o da LIPNet, considerou-se o alcance do número máximo de iterações.

Os experimentos foram realizados considerando os seguintes arranjos experimentais descritos na Tabela 3. Nesta, os valores indicados para a configuração da LIPNet correspondem, nesta ordem, a: tamanho do campo receptivo 1, *overlap* 1, tamanho do campo receptivo 2, *overlap* 2. Considerou-se o formato quadrado dos campos receptivos. Esta variação na configuração da LIPNet foi aplicada para avaliar o efeito do tamanho dos campos receptivos nos experimentos.

A Tabela 3 também mostra que as imagens geradas pelo modelo de visualização de imagens proposto neste trabalho foram formadas de duas formas: considerando apenas uma imagem da base de teste ou utilizando um certo número de imagens aleatórias da base de teste.

Como se pode notar, os experimentos realizados não avaliaram a profundidade da rede. Isto se deve ao fato de as imagens utilizadas serem pequenas, não necessitando da adição de múltiplas camadas. Assim, considerou-se apenas as duas camadas piramidais usualmente aplicadas por (FERNANDES, 2013).

Os códigos utilizados foram desenvolvidos em linguagem JAVA. Como ferramentas

Tabela 3 – Arranjos experimentais.

[Fonte: elaboração própria]

Bases	Variáveis	Valores
AR/CBCL	Fator Iterações PSO Iterações LIPNet AUC	0,1, 0,2 e 0,5 10, 100, 500, 1500 e 2500 10, 50, 100, 250, 500 Média de 10 simulações
CBCL	Configuração LIPNet Modelo de visualização	(4,1,3,0), (9,4,3,2) 1 imagem e média de 10 imagens
AR	Configuração LIPNet Modelo de visualização	(10,4,3,0), (14,1,3,2) 1 imagem e média de 300 imagens

auxiliares cita-se a IDE Eclipse e o Matlab, usado para realizar o pré-processamento nas imagens das bases de dados.

4.2 CBCL

Considerando primeiramente a AUC, os melhores valores encontrados estão exibidos na Tabela 4. Eles mostram que a média alcançada pelo modelo não-negativo não ultrapassa a média obtida pelo modelo tradicional.

Tabela 4 – Valores da AUC encontrados nos experimentos realizados com a CBCL.

[Fonte: elaboração própria]

Base	Modelo	AUC	Desvio	Pa- drão
CBCL	Tradicional	0,849	0,01	
	Não-negativo	0,823	0,04	

Por outro lado, as imagens geradas pelo modelo de visualização do aprendizado da rede demonstra que o modelo não-negativo forma imagens mais similares a faces em relação ao modelo tradicional, como mostra a Figura 11. Este resultado é consistente com o fato de a não-negatividade permitir apenas combinações aditivas para reconstruir imagens. Assim, os padrões aprendidos pelos neurônios ocultos da RNA não-negativa proporcionam uma interpretação mais fácil.

A comparação entre os resultados obtidos com as variações do *fator* da potenciação demonstra que, ao utilizar o valor 0,1, há uma formação de imagens mais nítidas, deixando a face mais visível. Esse fenômeno já era esperado pois, quanto menor o fator, mais detalhes se tornam visíveis. Esta comparação pode ser visualizada na Figura 12.

O progresso realizado durante o aprendizado dos padrões pode ser notado através da variação do número de iterações. A Figura 13 demonstra que, ao longo das iterações do PSO, a LIPNet não-negativa evolui claramente no aprendizado. As regiões dos olhos,

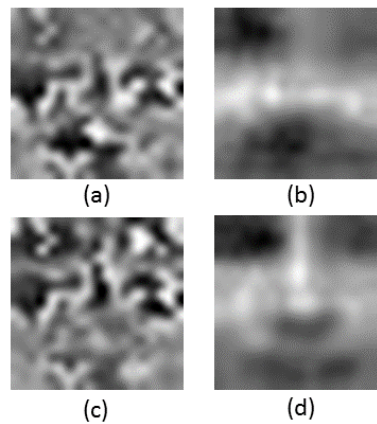


Figura 11 – (a) e (c) correspondem às representações de 300 imagens aleatórias do banco de teste da CBCL geradas pelo modelo de visualização do aprendizado da rede para a LIPNet tradicional de não-face e face, respectivamente; (b) e (d) correspondem às representações de 300 imagens aleatórias do banco de teste da CBCL geradas pelo modelo de visualização do aprendizado da rede para a LIPNet não-negativa de não-face e face, respectivamente. A configuração utilizada nestes resultados foi a $(4,1,3,0)$, o fator = 0,1, 2500 iterações do PSO e 500 iterações da LIPNet.

[Fonte: elaboração própria]

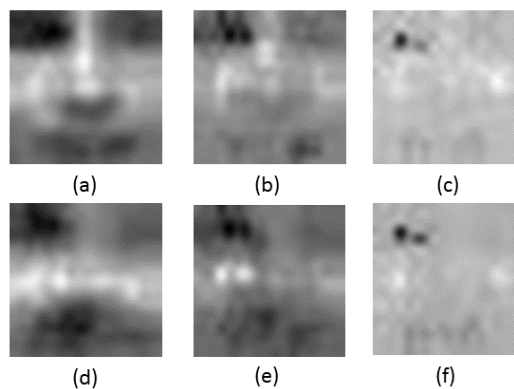


Figura 12 – (a), (b) e (c) são representações de 300 imagens aleatórias de faces do banco de teste da CBCL geradas pelo modelo de visualização do aprendizado da rede para a LIPNet com restrição correspondentes aos fatores 0,1, 0,2 e 0,5, respectivamente; (d), (e) e (f) são representações de 300 imagens aleatórias de não-faces do banco de teste da CBCL geradas pelo modelo de visualização do aprendizado da rede para a LIPNet com restrição correspondentes aos fatores 0,1, 0,2 e 0,5, respectivamente. A configuração utilizada nestes resultados foi a $(4,1,3,0)$ e 2500 iterações do PSO foram consideradas.

[Fonte: elaboração própria]

da boca e do nariz vão ficando mais definidas à medida que o número de iterações cresce. Aos poucos, a rede aprende quais pixels são mais importantes para a identificação de uma face. Já a Figura 14 mostra que a evolução do aprendizado não fica visível no modelo tradicional por este ser opaco.

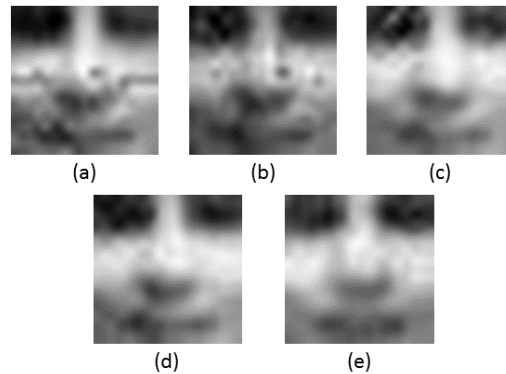


Figura 13 – (a), (b), (c), (d) e (e) são imagens que correspondem a representação de 300 imagens aleatórias de face do banco de teste da CBCL geradas pelo modelo de visualização do aprendizado da rede para a LIPNet com restrição correspondentes aos valores de iterações do PSO 10, 100, 500, 1500 e 2500, respectivamente. A configuração utilizada nestes resultados foi a (9,4,3,2) e o fator=0,1.

[Fonte: elaboração própria]

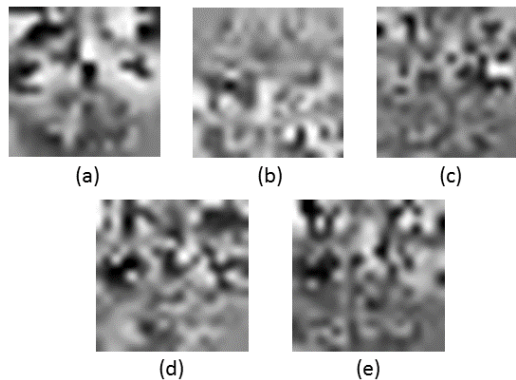


Figura 14 – (a), (b), (c), (d) e (e) correspondem à representação de 300 imagens aleatórias de face do banco de teste da CBCL geradas pelo modelo de visualização do aprendizado da rede para a LIPNet sem restrição correspondentes aos valores de iterações da rede 10, 50, 100, 250 e 500, respectivamente. A configuração utilizada nestes resultados foi a (4,1,3,0) e o fator=0,1.

[Fonte: elaboração própria]

A Figura 15 demonstra que ao utilizar um tamanho de campo receptivo maior, as imagens geradas ficaram mais nítidas. Isto pode se dever ao fato de que, quanto mais

conexões entre os neurônios existirem, melhor a rede vai interpretar um padrão. Pois, dessa forma, seria possível extrair mais informações de uma camada para outra, evoluindo no aprendizado. Outros testes variando-se a configuração dos campos receptivos podem ser aplicados para validar esta hipótese.

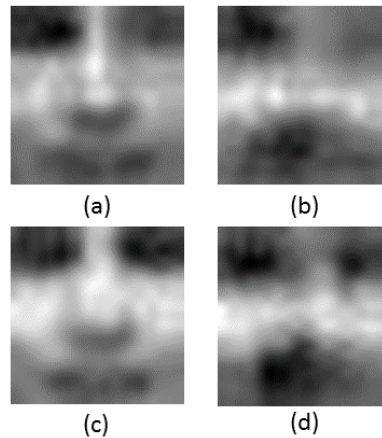


Figura 15 – (a) e (b) correspondem à representação de 300 imagens aleatórias de faces da base de teste da CBCL considerando a configuração (4,1,3,0) geradas pelo modelo de visualização do aprendizado da rede para a LIPNet com restrição; (c) e (d) correspondem à representação de 300 imagens aleatórias de faces da base de teste da CBCL considerando a configuração (9,4,3,2) geradas pelo modelo de visualização do aprendizado da rede para a LIPNet com restrição. O fator utilizado para estas imagens foi 0,1 e 2500 iterações do PSO foram consideradas.

[Fonte: elaboração própria]

4.3 AR

Como a Tabela 5 mostra, os resultados da AUC para a base de dados AR também foram inferiores utilizando a LIPNet não-negativa. Em contrapartida, assim como nos experimentos da CBCL, a interpretação do modelo não-negativo se mostrou superior. Nestes casos, os pesos são mais esparsos. Já a RNA tradicional gera imagens com pesos menos localizados, o que dificulta a visualização do padrão aprendido.

Tabela 5 – Valores da AUC encontrados nos experimentos realizados com a AR.

[Fonte: elaboração própria]

Base	Modelo	AUC	Desvio	Pa- drão
AR	Tradicional	0,964	0.006	
	Não-negativo	0,939	0.005	



Figura 16 – (a) e (e) mostram as 10 imagens do banco de teste da AR de cada gênero que foram consideradas; (b) e (f) são imagens que correspondem à representação dessas imagens geradas pelo modelo de visualização do aprendizado da rede para a LIPNet tradicional; (c) e (g) são imagens que correspondem à representação dessas imagens geradas pelo modelo de visualização do aprendizado da rede para a LIPNet não-negativa; por fim, (d) e (h) são as médias das 10 imagens. A configuração utilizada nestes resultados foi a $(14,1,3,2)$, o fator = 0,1, 2500 iterações do PSO e 500 iterações da LIPNet.

[Fonte: elaboração própria]

A Figura 16 também demonstra que as imagens geradas pelo modelo não-negativo, ao considerar 10 imagens da base de teste da AR, aparentam ser uma mulher e um homem específicos, ao invés de uma representação mais geral. Porém, essas imagens criadas não pertencem a pessoas existentes na base de teste. Como pode ser visualizado também na Figura 17, a LIPNet não-negativa criou *templates* do que ela entendeu como mulher e homem a partir das imagens da base AR fornecidas.

Como se pode notar ao comparar as imagens geradas pelo modelo não-negativo com as imagens médias, elas não são idênticas. A imagem média das 5 mulheres exibida na Figura 17, por exemplo, mostra um leve contorno de óculos, visto que há 2 mulheres usando esse acessório no conjunto considerado. Já na Figura 16, a imagem formada pelo modelo de visualização para os 10 homens possui um pouco mais cabelo que a imagem média.

Os resultados obtidos na AR em relação ao valor do *fator* foram equivalentes aos alcançados pela CBCL. A Figura 18 apresenta a comparação entre os três valores usados

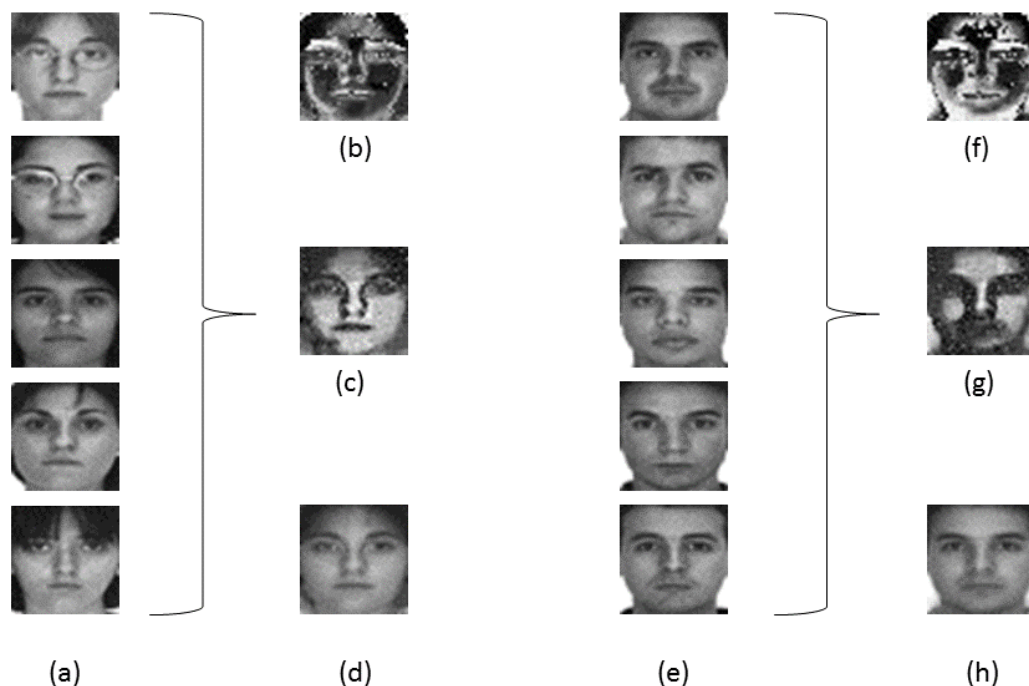


Figura 17 – (a) e (e) mostram as 5 imagens do banco de teste da AR de cada gênero que foram consideradas; (b) e (f) são imagens que correspondem à representação dessas imagens geradas pelo modelo de visualização do aprendizado da rede para a LIPNet tradicional; (c) e (g) são imagens que correspondem à representação dessas imagens geradas pelo modelo de visualização do aprendizado da rede para a LIPNet não-negativa; por fim, (d) e (h) são as médias das 5 imagens. A configuração utilizada nestes resultados foi a $(14,1,3,2)$, o fator = 0,1, 2500 iterações do PSO e 500 iterações da LIPNet.

[Fonte: elaboração própria]

para o *fator*, na qual o valor 0,1 obteve as imagens mais visíveis.

Na Figura 19, vê-se as imagens geradas para cada número de iterações do PSO testado. Neste caso, o progresso do aprendizado se mostrou muito mais discreto. A Figura 20 mostra a evolução das iterações da LIPNet no modelo tradicional. Estas continuam sendo imagens que não representam claramente uma face.

Por fim, a Figura 21 mostra que na comparação entre os tamanhos dos campos receptivos realizada na AR não apresenta grandes diferenças entre as duas configurações utilizadas, provavelmente pela quantidade de detalhes já existentes nas imagens.

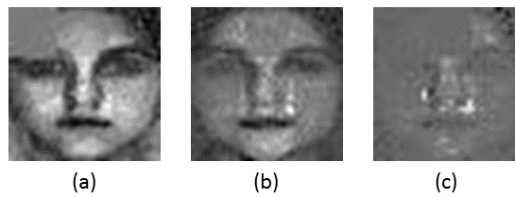


Figura 18 – (a), (b) e (c) são imagens correspondente a 10 mulheres do banco de teste da AR geradas pelo modelo de visualização do aprendizado da rede para a LIPNet com restrição correspondentes aos fatores 0,1, 0,2 e 0,5, respectivamente. A configuração utilizada nestes resultados foi a (14,1,3,2) e 2500 iterações do PSO foram consideradas.

[Fonte: elaboração própria]

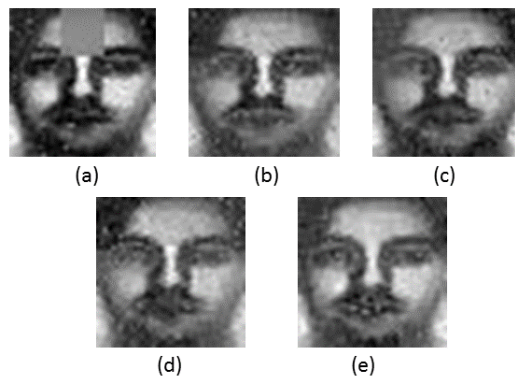


Figura 19 – (a), (b), (c), (d) e (e) são imagens que correspondem a 10 homens do banco de teste da AR geradas pelo modelo de visualização do aprendizado da rede para a LIPNet com restrição correspondentes aos valores de iterações do PSO 10, 100, 500, 1500 e 2500, respectivamente. A configuração utilizada nestes resultados foi a (14,1,3,2) e o fator=0,1.

[Fonte: elaboração própria]

4.4 Principais considerações

Nos dois casos, as AUCs encontradas não ultrapassam o modelo sem restrição. Tal resultado também pôde ser encontrado no trabalho de Chorowski e Zurada ([CHOROWSKI; ZURADA, 2015](#)), onde as taxas foram próximas mas não superaram o modelo que permite valores negativos. Eles defendem que em certas situações uma melhor visão sobre os dados supera os benefícios de um classificador com melhor acurácia mas opaco.

Além disso, as imagens formadas pelo modelo de visualização do aprendizado da rede considerando a base AR expõe uma interpretação mais clara da LIPNet tradicional em comparação com aos resultados obtidos pela CBCL. Isto pode ser visualizado nas Figuras 11 e 16. Isto pode se dever ao fato de as imagens da CBCL serem bem menores

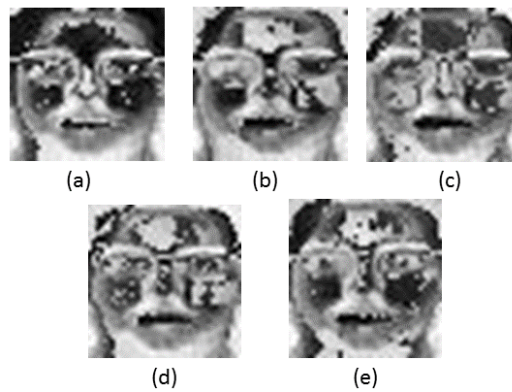


Figura 20 – (a), (b), (c), (d) e (e) são imagens que correspondem a 10 homens do banco de teste da AR geradas pelo modelo de visualização do aprendizado da rede para a LIPNet sem restrição correspondentes aos valores de iterações da rede 10, 50, 100, 250 e 500, respectivamente. A configuração utilizada nestes resultados foi a (14,1,3,2) e o fator=0,1.

[Fonte: elaboração própria]

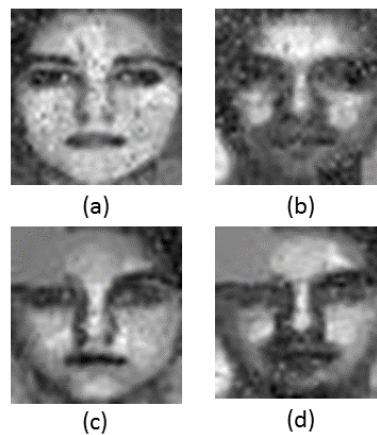


Figura 21 – (a) e (b) correspondem à representação de 10 imagens de mulheres e homens, respectivamente, da base de teste da AR considerando a configuração (4,1,3,0) geradas pelo modelo de visualização do aprendizado da rede para a LIPNet com restrição; (c) e (d) correspondem à representação de 10 imagens de mulheres e homens, respectivamente, da base de teste da AR considerando a configuração (14,1,3,2) geradas pelo modelo de visualização do aprendizado da rede para a LIPNet com restrição. O fator utilizado para estas imagens foi 0,1 e 2500 iterações do PSO foram consideradas.

[Fonte: elaboração própria]

que as da AR, sendo 19x19 e 40x40 os tamanhos destas, respectivamente. Além disso, as imagens da AR estão bem centralizadas. Assim, as imagens formadas a partir das imagens da AR possuem mais detalhes.

5 Considerações Finais

5.1 Conclusões

Neste trabalho foi realizado um estudo sobre a influência da restrição não-negativa no aprendizado da LIPNet. Assim, os pesos utilizados na RNA foram não-negativos e a função de ativação usada foi apenas a sigmóide-logística, que gera valores no intervalo $[0,1]$. Para aplicar a restrição de pesos não-negativos, foi utilizado o PSO. Além disso, um modelo de visualização do aprendizado da rede foi proposto com o objetivo de comparar a interpretação realizada pela LIPNet tradicional e a não-negativa.

Pôde-se concluir que embora não tenha havido uma melhoria nas taxas de classificação, a interpretabilidade da rede se mostrou superior na rede neural não-negativa. Este resultado é coerente com o fato de a restrição não-negativa permitir apenas combinações positivas, facilitando a identificação dos conceitos mais importantes que formam os padrões. Notou-se que a LIPNet com restrição conseguiu gerar imagens mais similares a faces humanas, localizando e separando de maneira proporcional as principais regiões da face: olhos, nariz, boca e bochechas. Com isso, mostra-se a capacidade da rede neural com restrição não-negativa de focar em partes mais simples e representativas para então analisar o todo, ou seja, realizar um aprendizado por partes.

Além disso, foi observado, nos resultados obtidos considerando a base AR, que a LIPNet não-negativa criou *templates* que representam as duas classes aprendidas: homem e mulher. Aplicações para esses *templates* podem ser encontradas após pesquisas futuras. Como exemplos de possíveis aplicações, cita-se tarefas de compressão e sintetização. Através de testes, pode-se concluir se esses *templates* podem ser usados para treinar a rede, ao invés de utilizar uma base de treino bem maior, diminuindo o custo computacional, por exemplo.

Por fim, ressalta-se que há casos nos quais é mais vantagem utilizar um classificador com menor acurácia, mas com uma maior interpretabilidade dos dados. Pode-se, por exemplo, utilizar de alguma forma o que a RNA aprendeu em um modelo não-negativo antes de fornecer os padrões de entrada para um outro classificador. Assim, é importante ter uma ferramenta que, não só realiza a classificação dos dados, como também gera novos conceitos e *insights* provenientes dos dados que podem ser aplicados em outras tarefas.

5.2 Trabalhos Futuros

Como trabalhos futuros propõe-se:

-
- Realizar testes com outras bases de dados para problemas de visão computacional como reconhecimento de expressões faciais e outros que não necessariamente envolvam faces;
 - Avaliar a curva de convergência do PSO dos modelos tradicional e não-negativo para cada experimento;
 - Realizar testes trocando a função sigmóide-logística por funções gaussianas, visto que essas também podem gerar apenas valores não-negativos;
 - Adicionar restrição de entradas positivas para analisar se há melhoras na interpretabilidade da rede e na taxa de classificação;
 - Buscar aplicações para os *templates* gerados pela LIPNet não-negativa;
 - Aprofundar os estudos na conexão existente entre a não-negatividade e o aprendizado por partes nas redes neurais.

Referências

- ABDULLAH, S. L. S. et al. Comparative study of random-pso and linear-pso algorithms. In: *Computer Information Science (ICCIS), 2012 International Conference on*. [S.l.: s.n.], 2012. v. 1, p. 409–413.
- BENGIO, Y. et al. *Scaling learning algorithms towards AI*. 2007.
- BEZERRA, B.; ZANCHETTIN, C.; ANDRADE, V. de. A mdrnn-svm hybrid model for cursive offline handwriting recognition. In: *Artificial Neural Networks and Machine Learning ICANN 2012*. [S.l.]: Springer Berlin Heidelberg, 2012, (Lecture Notes in Computer Science, v. 7553). p. 246–254. ISBN 978-3-642-33265-4.
- BISHOP, C. M. *Neural Networks for Pattern Recognition*. New York, NY, USA: Oxford University Press, Inc., 1995. ISBN 0198538642.
- BRADLEY, A. P. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition*, v. 30, n. 7, p. 1145 – 1159, 1997. ISSN 0031-3203.
- CHEN, J. et al. Nonnegative least-mean-square algorithm. *Signal Processing, IEEE Transactions on*, v. 59, n. 11, p. 5225–5235, Nov 2011. ISSN 1053-587X.
- CHEN, Y. et al. Discriminative multi-modality non-negative sparse graph model for action recognition. In: *Visual Communications and Image Processing Conference, 2014 IEEE*. [S.l.: s.n.], 2014. p. 53–56.
- CHOROWSKI, J.; ZURADA, J. Learning understandable neural networks with nonnegative weight constraints. *Neural Networks and Learning Systems, IEEE Transactions on*, v. 26, n. 1, p. 62–69, Jan 2015. ISSN 2162-237X.
- CLERC, M.; KENNEDY, J. The particle swarm - explosion, stability, and convergence in a multidimensional complex space. *Evolutionary Computation, IEEE Transactions on*, v. 6, n. 1, p. 58–73, Feb 2002. ISSN 1089-778X.
- DANIELS, H.; KAMP, B. Application of mlp networks to bond rating and house pricing. *Neural Computing & Applications*, Springer-Verlag London Limited, v. 8, n. 3, p. 226–234, 1999. ISSN 0941-0643.
- DANIELS, H.; VELIKOVA, M. Monotone and partially monotone neural networks. *Neural Networks, IEEE Transactions on*, v. 21, n. 6, p. 906–917, June 2010. ISSN 1045-9227.
- ENGELBRECHT, A. P. *Computational Intelligence: An Introduction*. 2nd. ed. [S.l.]: Wiley Publishing, 2007. ISBN 0470035617.
- FENG, M.; PAN, H. A modified pso algorithm based on cache replacement algorithm. In: *Computational Intelligence and Security (CIS), 2014 Tenth International Conference on*. [S.l.: s.n.], 2014. p. 558–562.

- FERNANDES, B.; CAVALCANTI, G.; REN, T. Autoassociative pyramidal neural network for face verification. In: *Neural Networks (IJCNN), The 2011 International Joint Conference on*. [S.l.: s.n.], 2011. p. 1612–1617. ISSN 2161-4393.
- FERNANDES, B. J. T. *Redes Neurais com Extração Implícita de Características para Reconhecimento de Padrões Visuais*. Tese (Doutorado) — Centro de Informática - CIn - UFPE, 2013.
- FERNANDES, B. T.; CAVALCANTI, G.; REN, T. I. Lateral inhibition pyramidal neural network for image classification. *Cybernetics, IEEE Transactions on*, v. 43, n. 6, p. 2082–2092, Dec 2013. ISSN 2168-2267.
- FUKUSHIMA, K.; MIYAKE, S.; ITO, T. Neocognitron: A neural network model for a mechanism of visual pattern recognition. *Systems, Man and Cybernetics, IEEE Transactions on*, SMC-13, n. 5, p. 826–834, Sept 1983. ISSN 0018-9472.
- FYFE, C. A neural network for pca and beyond. *Neural Processing Letters*, Kluwer Academic Publishers, v. 6, n. 1-2, p. 33–41, 1997. ISSN 1370-4621.
- GE, X.; IWATA, S. Learning the parts of objects by auto-association. *Neural Networks*, v. 15, n. 2, p. 285 – 295, 2002. ISSN 0893-6080.
- GHOSH, K.; PAL, S. Some insights into brightness perception of images in the light of a new computational model of figure-ground segregation. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, n. 4, p. 758–766, July 2010. ISSN 1083-4427.
- GONZALEZ, R.; WOODS, R. *Processamento Digital De Imagens*. [S.l.]: Addison Wesley Bra, 2011. ISBN 9788576054016.
- GOUVEIA, F. M. et al. Handwriting recognition system for mobile accessibility to the visually impaired people. In: *Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on*. [S.l.: s.n.], 2014. p. 3918–3923.
- GRIGORESCU, C.; PETKOV, N.; WESTENBERG, M. Contour detection based on nonclassical receptive field inhibition. *Image Processing, IEEE Transactions on*, v. 12, n. 7, p. 729–739, July 2003. ISSN 1057-7149.
- HE, R. et al. Nonnegative sparse coding for discriminative semi-supervised learning. In: *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. [S.l.: s.n.], 2011. p. 2849–2856. ISSN 1063-6919.
- HEISELE, B.; POGGIO, T.; PONTIL., M. *Face detection in still gray images*. [S.l.], 2000.
- KENNEDY, J.; EBERHART, R. Particle swarm optimization. In: *Neural Networks, 1995. Proceedings., IEEE International Conference on*. [S.l.: s.n.], 1995. v. 4, p. 1942–1948 vol.4.
- LECUN, Y. et al. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, v. 1, n. 4, p. 541–551, Dec 1989. ISSN 0899-7667.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, Nov 1998. ISSN 0018-9219.

- LEMME, A.; REINHART, R. F.; STEIL, J. J. Efficient online learning of a non-negative sparse autoencoder. In: D-FACTO. *ESANN*. Bruges: d-facto, 2010. doi:10.1016/j.humov.2010.04.001, p. 1–6.
- LEVINE, M. W.; SHEFNER, J. M. *Fundamentals of Sensation and Perception*. [S.l.]: Brooks/Cole Publishing Company, 1991.
- LUAN, Q. et al. Video summarization based on nonnegative linear reconstruction. In: *Multimedia and Expo (ICME), 2014 IEEE International Conference on*. [S.l.: s.n.], 2014. p. 1–6.
- MA, J. The stability of asymmetric hopfield networks with nonnegative weights. In: *Theoretical Aspects of Neural Computation: A Multidisciplinary Perspective, Proceedings of Hong Kong International Workshop (TANC'97)*. [S.l.: s.n.], 1997.
- MAO, Z.-H.; MASSAQUOI, S. Dynamics of winner-take-all competition in recurrent neural networks with lateral inhibition. *Neural Networks, IEEE Transactions on*, v. 18, n. 1, p. 55–69, Jan 2007. ISSN 1045-9227.
- MARTINEZ, A. M.; BENAVENTE, R. *The AR Face Database*. [S.l.], 1998.
- NOLKER, C.; RITTER, H. Visual recognition of continuous hand postures. *Neural Networks, IEEE Transactions on*, v. 13, n. 4, p. 983–994, Jul 2002. ISSN 1045-9227.
- PARK, H.-S.; PEDRYCZ, W.; OH, S.-K. Granular neural networks and their development through context-based clustering and adjustable dimensionality of receptive fields. *Neural Networks, IEEE Transactions on*, v. 20, n. 10, p. 1604–1616, Oct 2009. ISSN 1045-9227.
- PEREZ, C. et al. Genetic design of biologically inspired receptive fields for neural pattern recognition. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, v. 33, n. 2, p. 258–270, Apr 2003. ISSN 1083-4419.
- PHUNG, S.; BOUZERDOUM, A. A pyramidal neural network for visual pattern recognition. *Neural Networks, IEEE Transactions on*, v. 18, n. 2, p. 329–343, March 2007. ISSN 1045-9227.
- PHUNG, S.; BOUZERDOUM, A. A pyramidal neural network for visual pattern recognition. *Neural Networks, IEEE Transactions on*, v. 18, n. 2, p. 329–343, March 2007. ISSN 1045-9227.
- PLUMBLEY, M. D.; OJA, E. A "nonnegative pca" algorithm for independent component analysis. *Trans. Neur. Netw.*, IEEE Press, Piscataway, NJ, USA, v. 15, n. 1, p. 66–76, jan. 2004. ISSN 1045-9227.
- RIEDMILLER, M.; BRAUN, H. A direct adaptive method for faster backpropagation learning: the rprop algorithm. In: *Neural Networks, 1993., IEEE International Conference on*. [S.l.: s.n.], 1993. p. 586–591 vol.1.
- RIZZOLATTI, G.; CAMARDA, R. Inhibition of visual responses of single units in the cat visual area of the lateral suprasylvian gyrus (clare-bishop area) by the introduction of a second visual stimulus. *Brain Research*, v. 88, n. 2, p. 357 – 361, 1975. ISSN 0006-8993.

- SHARIFARA, A.; RAHIM, M. M.; ANISI, Y. A general review of human face detection including a study of neural networks and haar feature-based cascade classifier in face detection. In: *Biometrics and Security Technologies (ISBAST), 2014 International Symposium on*. [S.l.: s.n.], 2014. p. 73–78.
- SOARES, A. M.; FERNANDES, B. J. T.; BASTOS-FILHO, C. J. A. Lateral inhibition pyramidal neural networks designed by particle swarm optimization. In: *Artificial Neural Networks and Machine Learning ICANN 2014*. [S.l.]: Springer International Publishing, 2014, (Lecture Notes in Computer Science, v. 8681). p. 667–674. ISBN 978-3-319-11178-0.
- SONG, L. et al. Prediction for chaotic time series of optimized bp neural network based on modified pso. In: *Control and Decision Conference (2014 CCDC), The 26th Chinese*. [S.l.: s.n.], 2014. p. 697–702.
- VU, X. T. et al. A new stochastic optimization algorithm to decompose large nonnegative tensors. *Signal Processing Letters, IEEE*, v. 22, n. 10, p. 1713–1717, Oct 2015. ISSN 1070-9908.
- XIUJUN, Z.; CHANG, L. Generalized discriminant orthogonal nonnegative tensor factorization for facial expression recognition. *The Scientific World Journal*, v. 2014, 2014.
- XU, L. Least mean square error reconstruction principle for self-organizing neural-nets. *Neural Networks*, v. 6, n. 5, p. 627 – 648, 1993. ISSN 0893-6080.
- YASUDA, M. Effective mean-field inference method for nonnegative boltzmann machines. In: *Pattern Recognition (ICPR), 2014 22nd International Conference on*. [S.l.: s.n.], 2014. p. 3600–3605. ISSN 1051-4651.
- ZASS, R.; SHASHUA, A. Nonnegative sparse pca. In: *In Neural Information Processing Systems*. [S.l.: s.n.], 2007.
- ZHANG, Y.; YU, B. Face recognition using combined non-negative principal component analysis and linear discriminant analysis. In: *Image and Signal Processing (CISP), 2013 6th International Congress on*. [S.l.: s.n.], 2013. v. 2, p. 758–762.
- ZHANG, Z.; LI, H.; HUANG, L. Topicdsdr: Combining topic decomposition and data reconstruction for summarization. In: WANG, J. et al. (Ed.). *Web-Age Information Management*. [S.l.]: Springer Berlin Heidelberg, 2013, (Lecture Notes in Computer Science, v. 7923). p. 338–350. ISBN 978-3-642-38561-2.
- ZHUANG, L. et al. Non-negative low rank and sparse graph for semi-supervised learning. In: *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. [S.l.: s.n.], 2012. p. 2328–2335. ISSN 1063-6919.