

بهبود مانیفولد حالات و تصاویر مجازی با به کارگیری شبکه‌های عصبی دوسویه در بازشناسی چهره با یک تصویر از هر فرد

فاطمه عبدالعلی و سیدعلی سیدصالحی

تاکنون روش‌های مبتنی بر شبکه‌های عصبی بسیاری برای حل مسئله بازشناسی چهره پیشنهاد شده‌اند. یکی از نخستین شبکه‌های عصبی مورد استفاده در بازشناسی چهره، شبکه تک‌لایه WISARD [۲] بود که شامل یک شبکه جداگانه برای هر شخص است. لین و همکارانش از یک شبکه عصبی بر مبنای احتمال^۴ استفاده نمودند که در واقع از یک شبکه برای هر تصویر چهره استفاده می‌شود و مستلزم وجود چند تصویر از هر فرد در تعلیم شبکه می‌باشد [۳]. لورنس و همکارانش نیز از یک شبکه عصبی هیبریدی استفاده نمودند که نمونه‌برداری محلی از تصویر، نگاشت خودسازمانده^۵ و شبکه عصبی پیچشی^۶ را ترکیب می‌نمود [۴]. در این رویکرد، از نگاشت خودسازمانده جهت کاهش بعد استفاده می‌شود که فضای تصویر با بعد بالا به فضای گسسته با بعد پایین‌تر نگاشت می‌شود و با گره‌ها نمایش داده می‌شود. به هر گره مجموعه‌ای از n وزن تخصیص می‌یابد که m بعد زیر تصویر است. در حین تعلیم، برای هر زیر تصویر تعلیم، واحد با بیشترین انطباق^۷ انتخاب می‌شود و واحد با بیشترین انطباق و گره‌های اطراف آن با داده‌های تعلیم تنظیم می‌شوند که این فرآیند توسط تابع همسایگی کنترل می‌شود و با گذشت زمان تابع همسایگی به تدریج به سوی صفر کاهش می‌یابد. در مرحله استخراج ویژگی و طبقه‌بندی نیز از یک شبکه پیچشی استفاده می‌شود. همچنین وونگ و همکارانش از یک شبکه عصبی با توابع پایه شعاعی به همراه روش تعلیم جدیدی بر مبنای الگوریتم حداقل مربعات خطای متعامد و تنظیم شده^۸ استفاده نمودند که منجر به کاهش پیچیدگی محاسبات و افزایش مقاوم بودن مدل در کاربردهای واقعی شد [۵].

در [۶] و [۷] به منظور بازشناسی چهره با یک تصویر از هر فرد، از رویکرد جداسازی اطلاعات فرد از حالت و تخمین مانیفولدهای مربوطه به کمک شبکه عصبی تحلیل‌گر غیر خطی چهره استفاده شده است. با به کارگیری مانیفولدهای تخمین زده شده و با هدف افزایش تعداد نمونه‌های تعلیم، تصاویر مجازی از تنها تصویر موجود از هر فرد در پایگاه داده تولید شده است. همچنین یک روش تعلیم بر مبنای خوشه‌بندی بدون سرپرستی اطلاعات فرد و حالت ارائه شده است.

باید تلاش‌ها در این زمینه با الهام از سامانه بازشناسی چهره مغز انسان ادامه یابند تا به بازشناسی مقاوم به تنوعات الگوها دست یابیم. برخی تحقیقات اخیر نشان داده‌اند که عملکرد مقاوم سامانه ادراک انسان در بازشناسی گفتار، چهره و غیره از شیوه خاص پردازش سیگنال که در آن انجام می‌شود، ناشی می‌گردد. از مشخصه‌های این شیوه پردازش که به قشر مخ نسبت داده می‌شود، می‌توان به نکات زیر اشاره نمود:

چکیده: در این مقاله به منظور توسعه عملکرد مدل‌های شبکه عصبی در بازشناسی چهره با یک تصویر از هر فرد، یک ساختار شبکه عصبی دوسویه با الهام از نئوکورتکس مغز انسان ارائه شده است. در ساختار پیشنهادی همانند نئوکورتکس در ابتدا طی یک مرحله پردازش از پایین به بالا، یک تفسیر زمخت از ورودی صورت می‌گیرد، سپس در مرحله بعد نتایج بازشناسی اولیه ضمن عبور از یک شبکه عصبی معکوس یالایش می‌شوند. از این مدل جهت جداسازی غیر خطی اطلاعات فرد از حالت و تخمین مانیفولدهای اطلاعات فرد و حالت استفاده شده است. به منظور افزایش تعداد نمونه‌های تعلیم در شبکه طبقه‌بندی کننده با استفاده از مانیفولدهای تخمین زده شده، تصاویر مجازی چهره‌های نرمال موجود در پایگاه داده تست تولید شده است. با تعلیم شبکه طبقه‌بندی کننده توسط تصاویر مجازی حاصل از تعلیم شبکه دوسویه، درصد صحت بازشناسی ۸۵/۴۵٪ روی دادگان تست حاصل شده که در مقایسه با تولید تصاویر مجازی با استفاده از روش خوشه‌بندی بدون سرپرستی اطلاعات افراد و حالات دارای بهبود ۱/۸۲٪ می‌باشد.

کلید واژه: بازشناسی چهره، یک تصویر از هر فرد، شبکه عصبی بازگشتی، پویایی جاذب، یادگیری مانیفولد، تصاویر مجازی، پالایش غیر خطی.

۱- مقدمه

از آنجایی که بازشناسی چهره به دلیل سهولت استفاده و کاربرپسند بودن، یکی از روش‌های پرکاربرد در زیست‌سنجی^۱ محسوب می‌شود، سامانه‌های بازشناسی چهره باید دارای قابلیت تشخیص چهره در محیط کنترل نشده و شرایط دلخواه باشند. این مسئله یعنی عمومی بودن شرایط و محیط، سبب به وجود آمدن چالش‌های جدی در تکنیک‌های بازشناسی چهره شده است، چرا که سامانه بازشناسی باید نسبت به تنوعات ناشی از تغییر حالت یا شرایط عکس‌برداری مقاوم باشد. رویکردهای بازشناسی چهره در مقابله با تنوعات را می‌توان به دو دسته تقسیم نمود. در دسته اول، پایگاه داده شامل تصاویر متعددی از هر فرد برای تعلیم است اما در دسته دوم، از هر فرد تنها یک تصویر برای تعلیم در پایگاه داده موجود است. البته این نحوه دسته‌بندی منحصر به فرد نمی‌باشد و بر مبنای معیارهای دیگری از جمله پیچیدگی محاسبات، مبتنی بر ویژگی^۲ یا ظاهر^۳ و غیره نیز قابل انجام است [۱].

این مقاله در تاریخ ۱۰ آبان ماه ۱۳۸۹ دریافت و در تاریخ ۳۱ اردیبهشت ماه ۱۳۹۰ بازنگری شد.

فاطمه عبدالعلی، دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر، تهران، (email: fabdolali@aut.ac.ir)

سیدعلی سیدصالحی، دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر، تهران، (email: ssalehi@aut.ac.ir)

4. PDBNN

5. Self-Organizing Map

6. Convolutional Neural Network

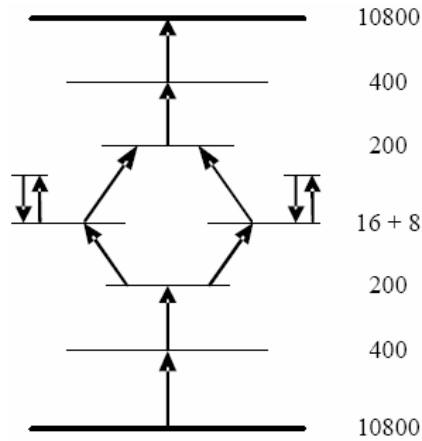
7. Best Matching Unit

8. Regularized Orthogonal Least Square

1. Biometrics

2. Feature-Based

3. Appearance-Based



شکل ۲: ساختار شبکه عصبی تحلیل‌گر جهت خوشه‌بندی ویژگی‌های افراد و حالات در لایه گلوگاه.

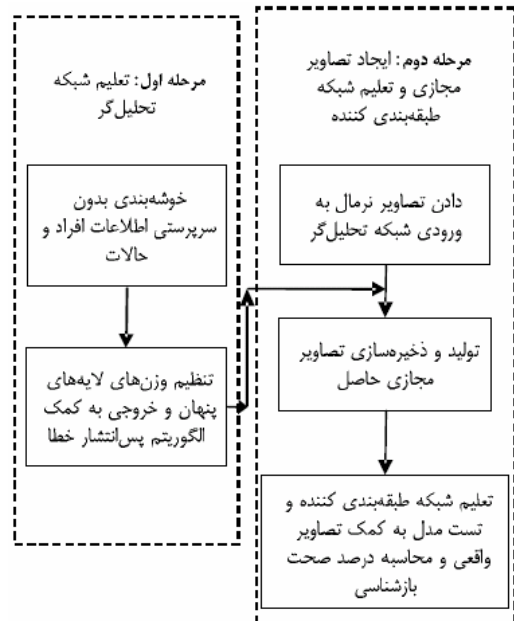
می‌پردازیم. در بخش سوم، این مدل با الهام از مدل نئوکورتکس و با به کارگیری یک شبکه معکوس جهت پردازش دوسویه اطلاعات توسعه داده می‌شود. در بخش چهارم نتایج شبیه‌سازی و نحوه ارزیابی کمی عملکرد شبکه ارائه می‌شوند. در نهایت بخش پنجم شامل جمع‌بندی و نتیجه‌گیری است.

۲- جداسازی مبتنی بر خوشه‌بندی اطلاعات فرد و حالت

در این بخش به صورت اجمالی به بررسی روش جداسازی اطلاعات فرد از حالت مبتنی بر خوشه‌بندی بدون سرپرستی می‌پردازیم. بلوک دیاگرام سامانه بازشناسی چهره مبتنی بر خوشه‌بندی بدون سرپرستی اطلاعات افراد و حالات در شکل ۱ نمایش داده شده است. از آنجایی که در بازشناسی چهره با یک تصویر از هر فرد، با یک تصویر نمی‌توان به تعمیم لازم برای بازشناسی دست یافت، رویکرد پیشنهادی متداول تخمین مجازی حالات دیگر افراد است. از حالات مجازی تولید می‌توان جهت تعلیم مناسب مانیفولدهای حالات مختلف افراد به مدل استفاده کرد. ساختار شبکه عصبی پیشنهادی جهت جداسازی مؤلفه‌های فرد و حالت در شکل ۲ نشان داده شده است. لایه ورودی شبکه دارای تعداد نورون‌های برابر پیکسل‌های تصویر (۱۰۸۰۰ نورون) می‌باشد. لایه گلوگاه نیز شامل ۲۴ نورون است که ۸ نورون برای اطلاعات حالت و ۱۶ نورون برای اشخاص تخصیص یافته است. از این پس این شبکه را به اختصار شبکه تحلیل‌گر می‌نامیم. در این شبکه نورون‌های لایه گلوگاه برای جداسازی مؤلفه‌های فرد از حالت به دو بخش تقسیم شده است. در بخش مؤلفه‌های فرد، در هر بار تعلیم تصاویر یک شخص، متوسط مقادیر خروجی نورون‌های بخش خوشه‌بندی شخص توسط (۱) به دست آمده و از این کد به عنوان کد شخص در مرحله بعد تعلیم استفاده می‌شود. همچنین با میانگین‌گیری از کد ایجادشده برای حالت یکسان افراد مختلف و استفاده از آن در تعلیم‌های بعدی، حالات یکسان در یک خوشه قرار می‌گیرد [۶] و [۷]

$$M_{i+1} = (\gamma \times M_i) + (1 - \gamma) \times P \quad (1)$$

در این رابطه M نشان‌گر کد میانگین برای شخص یا حالت واحد، γ ضریب میزان تأثیر کد جدید (در اینجا ۰.۹۵ در نظر گرفته شده) و P کد تولیدشده برای تصویر جدید است. هر قدر γ بزرگ‌تر انتخاب شود، میزان تأثیر کدهای قبلی بیشتر خواهد بود. علت استفاده از روش تعلیم بدون سرپرستی این است که بدون سرپرست بودن، بخشی از لایه گلوگاه سبب ادغام اطلاعات مختلف در این لایه شده و اطلاعات مربوط به همه



شکل ۱: بلوک دیاگرام سامانه بازشناسی چهره پیشنهادی در [۶].

(۱) دوطرفه بودن پردازش سیگنال در شبکه‌های عصبی با اتصالات جلوسو و معکوس یا بازگشتی.

(۲) توانایی پردازش غیر خطی سیگنال‌ها و تجزیه و تحلیل مؤلفه‌های خطی و غیر خطی موجود در سیگنال ورودی و توانایی حذف اثر هر مورد در شرایط خاص مورد نظر. توانایی شبکه‌های عصبی بازگشتی در بهینه‌سازی و حرکت به سوی جاذب یا نقطه تعادل پایدار در این رابطه می‌تواند به عنوان ابزار مفیدی به کار گرفته شود [۸].

ساختارهای یک‌سویه شبکه‌های عصبی جلوسو در به کارگیری برای بازشناسی الگوها در برخی موارد نارسا هستند. از جمله این موارد بازشناسی الگوهایی است که تحت تأثیر نویز قرار گرفته‌اند و نیز بازشناسی الگوهایی که تحت تأثیر عوامل مختلف تغییر یافته‌اند که به آنها تنوعات الگوهای ورودی می‌گوییم. برخی تحلیل‌ها نشان می‌دهند که اضافه کردن اتصالات بازگشتی و نیز به کارگیری توأم یک شبکه جلوسو همراه با شبکه معکوس آن هنگامی که در شبکه نورون‌های غیر خطی داریم، قابلیت‌های پردازشی و بازشناسی شبکه را می‌تواند افزایش دهند [۹] و [۱۰]. از سوی دیگر شواهد موجود حاکی از پردازش دوسویه الگوهای مورد بازشناسی در مغز هستند [۱۱].

در این مقاله، در ادامه تلاش‌های صورت گرفته در [۶] و [۷]، با الهام از مدل نئوکورتکس، ساختار شبکه عصبی مبتنی بر خوشه‌بندی اطلاعات فرد و حالت توسعه داده می‌شود. با استفاده از ساختار پیشنهادی و پایگاه داده AUT که شامل تصاویر ۸۰ نفر در ۱۲ حالت و شرایط نورپردازی مختلف است، به جداسازی اطلاعات افراد از حالت پرداخته و سپس با ترکیب غیر خطی مؤلفه‌های حالات مختلف و مؤلفه‌های حالت نرمال اشخاص موجود در پایگاه داده تست (ZD)، به تولید تصاویر مجازی افراد موجود در پایگاه داده تست می‌پردازیم. در ادامه به منظور ارزیابی کیفیت تصاویر مجازی حاصل، از یک شبکه عصبی طبقه‌بندی کننده استفاده می‌شود. در این تحقیق تصاویر نرمال و تصاویر مجازی افراد پایگاه داده ZD به عنوان داده تعلیم و تصاویر واقعی این اشخاص در حالات مختلف به عنوان داده تست شبکه طبقه‌بندی کننده به کار رفته‌اند.

ساختار مقاله به این شرح است: در بخش دوم به معرفی دادگان مورد استفاده و مبانی روش تعلیم مبتنی بر خوشه‌بندی اطلاعات افراد از حالات



شکل ۳: تصاویر مربوط به شخص چهل و یکم و هفتاد و ششم از پایگاه داده AUC.

۳- ساختار شبکه عصبی پیشنهادی با الهام از مدل نئوکورتکس

هدف این مقاله پرداختن به شبیه‌سازی مدل مناسبی از شبکه‌های عصبی است که قادر باشد مانند انسان توسط یک سامانه بازشناسی، تصاویر ورودی با تنوعات گوناگون را به نحو بهینه و مناسب بازشناسی نماید. لازمه این کار اولاً تطبیقی بودن شبکه عصبی مورد استفاده برای بازشناسی است و ثانیاً عملکرد دوطرفه شبکه بازشناسی مانند مدل پیشنهادی Koerner برای قشر مخ می‌باشد [۱۲] تا [۱۴] که در آن عمل بازشناسی فقط طی یک مرحله انجام نمی‌شود بلکه طی مراحل متعددی از بازشناسی جلوسو و برگشت و دورزدن در شبکه انجام می‌شود و در هر مرحله آنچه که بازشناسی می‌شود، در مراحل بعدی برای تطابق مدل مورد استفاده قرار می‌گیرد.

ایجاد قدرت تطابق بهینه در مدل شبکه عصبی نسبت به نوع تصویر ورودی منجر به این می‌شود که مدل آنچه اشتراکات معرفتی موجود در دو دادگان است را به صورت مؤلفه‌های واحدی بیاموزد و در خود جا دهد و صرفاً در تفاوت‌ها خود را تطابق دهد. این کار را به بیان دیگر می‌توانیم مشابه خوشه‌بندی نمونه‌ها بدانیم که قدرت یادگیری و تعمیم را در سامانه‌های هوشمند ایجاد می‌کند. با این تفاوت که در اینجا مؤلفه‌های معرفتی و دانش موجود در شبکه است که خوشه‌بندی می‌شوند. با تحقق چنین مدلی، آنچه که مدل از یک دادگان با یک شرایط گوناگونی می‌آموزد، می‌تواند برای شرایط متنوع دیگر از آن استفاده کند [۱۵].

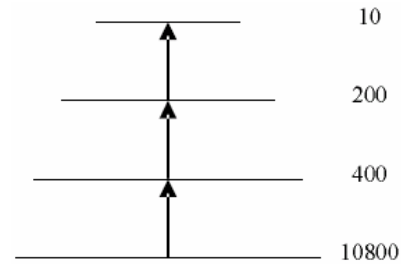
با این شیوه هم تعداد پارامترهای متغیر و تعلیم بیننده سامانه بازشناسی کاهش می‌یابد و هم دادگان متنوع می‌توانند با هم برای تعلیم سامانه به کار روند و یک بازشناسی با کیفیت بالا و مقاوم را در سامانه ایجاد نمایند. به نظر می‌رسد که اگر چنین نبود با توجه به حجم بالای تنوعات موجود در تصاویر چهره، انسان تا پایان عمر نیز قادر به بازشناسی خوب چهره نمی‌بود.

در تحقیقات اخیر، به‌منظور جبران تنوعات دو رویکرد کلی پیشنهاد شده است:

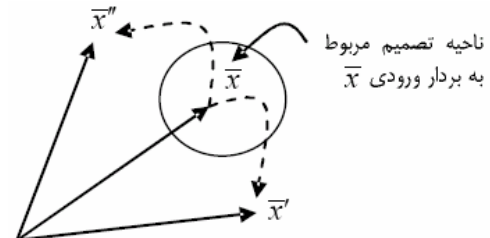
(۱) تطبیق مدل

(۲) تطبیق ورودی

در روش اول، مدل به‌گونه‌ای خود را تطبیق می‌دهد که ابرصفحات جابه‌جا شوند و بردار دچار تنوع‌شده، یعنی \vec{x}' یا \vec{x}'' در ناحیه تصمیم صحیح خود قرار بگیرد (شکل ۵). در واقع ناحیه تصمیم مربوطه حالت پویا



شکل ۴: ساختار شبکه عصبی طبقه‌بندی کننده.



شکل ۵: برداری که در اثر تنوعات در فضای بسیار بعدی حرکت کند، ممکن است که از ناحیه تصمیم مربوط به خود در یک سامانه بازشناسی غیر مقاوم نسبت به این تنوعات بیرون رود.

ویژگی‌ها از این بخش عبور می‌کنند، یعنی به‌طور کامل به هدف جداسازی اطلاعات در این حالت دست نمی‌یابیم. پس از تعلیم شبکه تحلیل‌گر توسط دادگان AUC و تولید تصاویر مجازی، از تصاویر مجازی حاصل جهت داده تعلیم شبکه طبقه‌بندی کننده استفاده می‌شود که ساختار آن در شکل ۳ نمایش داده شده است. این شبکه دارای دو لایه پنهان بوده، لایه ورودی شبکه دارای تعداد نورون‌های برابر پیکسل‌های تصویر (۱۰۸۰۰ نورون)، لایه پنهان اول دارای ۴۰۰ نورون و لایه پنهان دوم دارای ۲۰۰ نورون با تابع غیر خطی سیگموئید می‌باشد.

در این تحقیق بخشی از پایگاه داده AUC که در دانشکده برق دانشگاه صنعتی امیرکبیر تهیه شده است، به‌عنوان پایگاه داده عمومی مورد استفاده قرار گرفته است. این دادگان شامل ۹۶۰ تصویر از ۸۰ نفر می‌باشد. به‌ازای هر شخص ۱۲ تصویر موجود است که تغییرات این تصاویر شامل جهت چهره (نگاه روبه‌رو، نگاه به راست و چپ، نگاه به بالا و پایین)، ظاهر چهره با عینک، حالات احساسی چهره (لبخند، اخم، تعجب، چشمان بسته) و تفاوت در میزان روشنایی (نور از راست و نور از چپ) در نظر گرفته شده است. تصاویر به‌صورت سیاه و سفید با ۲۵۶ سطح خاکستری و دارای رزولوشن ۹۰×۱۲۰ می‌باشند. در شکل ۴ تصاویر مربوط به شخص چهل و یکم و هفتاد و ششم از پایگاه داده AUC نمایش داده شده است. پایگاه داده اصلی که آزمایش‌ها بر روی آن انجام می‌شود، شامل ۱۰ نفر و ۱۲ تصویر از هر فرد است که از این پس این پایگاه داده ZD نامیده می‌شود. این حالت‌ها مشابه حالت‌های پایگاه داده عمومی می‌باشد. این پایگاه داده در دانشکده مهندسی پزشکی دانشگاه صنعتی امیرکبیر تهیه شده و تمام تصاویر در پس‌زمینه یکسان تهیه شده و تنوع سنی نیز تا حد ممکن در آن رعایت شده است.

در [۶] نشان داده شده است که با ساختار پیشنهادی می‌توان به درصد صحت بازشناسی ۸۳/۶۳٪ روی دادگان تست دست یافت که نسبت به حالت تعلیم شبکه طبقه‌بندی کننده فقط با تصاویر نرمال دارای بهبود ۱۲/۷۳٪ است. جهت بهبود نحوه شکل‌گیری مؤلفه‌ها در شبکه تحلیل‌گر با الهام از شبکه آنالیز و سنتز دوسویه در مغز، از یک شبکه عصبی معکوس استفاده شده که جزئیات مربوط به ساختار نحوه تعلیم این شبکه در بخش بعدی به‌طور مفصل می‌آید.

می‌دهند، جابه‌جا می‌شوند. بنابراین باید دوباره این بردار جابه‌جا شده در ناحیه تصمیم مربوط به خود قرار بگیرد. به این ترتیب نواحی تصمیم در فضای بسیار بعدی پویا و دینامیک می‌شوند.

با به‌کارگیری یک شبکه واسط، خروجی شبکه معکوس به طرف دیگر داده می‌شود. هدف از افزودن شبکه معکوس به مدل همان‌گونه که قبلاً ذکر شد، گنجاندن خاصیت تطبیق در مدل است. به‌طور مثال می‌خواهیم که سامانه بازشناس، حالت را علی‌رغم تغییرات شخص تشخیص دهد.

در این مدل شبکه جلوسو همان مدل خوشه‌بندی بدون سرپرستی است که در ادامه آن یک شبکه عصبی معکوس قرار می‌دهیم که تنوعات فرد یا حالت به آن وارد می‌شوند، سپس با عبور از دو لایه پنهان این شبکه پردازش شده و خروجی این شبکه به لایه‌های پنهان شبکه جلوسو داده می‌شود. هدف از ارائه این مدل آن است که تنوعات فرد و حالت بعد از اینکه از دو لایه پردازش غیر خطی (طی شبکه معکوس) عبور داده شدند، بتوانند سطوح آستانه معادلات ابرصفحات تصمیم‌گیرنده اطلاعات فرد یا حالت را در فضای بسیار بعدی تغییر بدهند، ابرصفحات را جابه‌جا کرده و نواحی تصمیم را بهینه کنند. از آنجایی که خطای بین خروجی مطلوب و خروجی شبکه عصبی جلوسو پس‌انتشار شده و باعث اصلاح وزن‌ها می‌شود، بنابراین تغییر سطوح آستانه معادلات ابرصفحات تصمیم‌گیرنده اطلاعات فرد و حالت که از طریق وزن‌های اتصالات شبکه عصبی شکل می‌گیرند، در راستای بهبود بازشناسی خواهد بود. در (۲) تا (۴) روابط مربوط به لایه‌های این مدل آمده است (شکل ۶)

$$\bar{y}_{II} = f(\bar{s}_{II} W_{LA}) \quad (2)$$

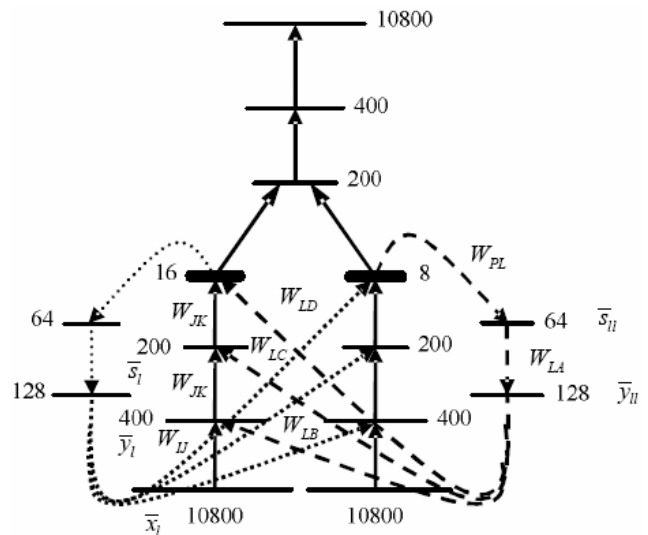
$$\bar{y}_I = f([\bar{x}_I \quad \bar{y}_{II}] \cdot \begin{bmatrix} W_{LI} \\ W_{LB} \end{bmatrix}) = f(\bar{x}_I W_{LI} + \bar{y}_{II} W_{LB}) \quad (3)$$

$$\bar{s}_I = f([\bar{y}_I \quad \bar{y}_{II}] \cdot \begin{bmatrix} W_{JK} \\ W_{LC} \end{bmatrix}) = f(\bar{y}_I W_{JK} + \bar{y}_{II} W_{LC}) \quad (4)$$

در این روابط W_{KP} و W_{JK} ، W_{LI} و W_{LA} وزن‌های شبکه مستقیم و W_{PL} ، W_{LD} و W_{LC} ، W_{LB} و W_{LA} وزن‌های شبکه معکوس می‌باشند. \bar{x}_I ، \bar{y}_I و \bar{s}_I به ترتیب بیانگر ورودی شبکه مستقیم، خروجی لایه پنهان اول و دوم می‌باشند. همچنین \bar{s}_{II} و \bar{y}_{II} خروجی لایه پنهان اول و دوم شبکه معکوس هستند. از روابط فوق مشخص می‌شود که در نگاهت لایه \bar{x} به لایه \bar{y} ، عبارت $\bar{y}_{II} W_{LB}$ از لایه \bar{y}_{II} به ورودی به لایه \bar{y}_I اضافه می‌شوند که در نگاهت لایه \bar{x} به لایه \bar{y} که متغیرهای نگاهت آن وزن‌های W_{LI} می‌باشند، اعداد ثابتی هستند و می‌توانند سطوح آستانه ابرصفحات تصمیم‌گیرنده این نگاهت را تغییر بدهند و در نتیجه ابرصفحات شکل‌دهنده نواحی تصمیم را به موازات خود در فضای بسیار بعدی جابه‌جا کنند. آموزش شبکه معکوس در ادامه شبکه جلوسو است، به این مفهوم که در ابتدای آموزش، وزن‌های اولیه شبکه معکوس مقادیر تصادفی کوچک و برای وزن‌های جلوسو، وزن‌های آموزش‌دیده در مرحله قبل قرار داده می‌شوند و اصلاح‌شدن یا نشدن این وزن‌ها تفاوت چندانی در نتیجه حاصل ایجاد نمی‌کند. بنابراین در این شبکه وزن‌های جلوسو که همان وزن‌های شبکه مرجع هستند، در طول آموزش اصلاح نمی‌شود.

۴- نتایج شبیه‌سازی

پس از تعلیم شبکه تحلیل‌گر، با دادن تصاویر نرمال پایگاه داده ZD به‌عنوان ورودی شبکه، تصاویر مجازی این افراد در ۱۲ حالت مختلف ایجاد می‌شود. در شکل ۷ تصاویر مجازی حاصل برای شخص هشتم از



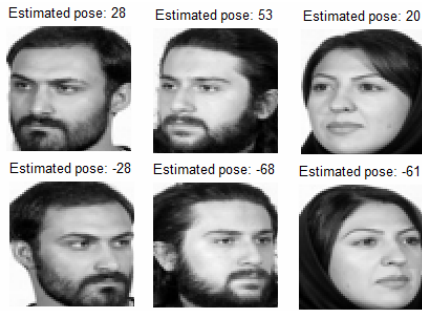
شکل ۶: ساختار شبکه عصبی پیشنهادی مبتنی بر [۹] جهت پویاکردن مدل‌های استخراج مؤلفه‌ها.

و دینامیک پیدا می‌کند و خود را به‌گونه‌ای تطبیق می‌دهد که تصمیم بهینه برای بردار ورودی گرفته شود. در روش دوم، ورودی خود را تطبیق می‌دهد، به این مفهوم که بردار ورودی حرکت می‌کند تا جایی که در نقطه بهینه تصمیم قرار گیرد. در هر دو حالت هدف این است که ورودی برای هر ناحیه تصمیم در بهترین نقطه آن ناحیه قرار گیرد.

در [۹] با الهام از شیوه بازشناسی ورودی‌های حسی در نئوکورتکس که به‌صورت حرکت‌های رفت و برگشتی می‌باشند، مدلی از بازشناسی گفتار در سطح آوا طراحی شده که قادر به مدل‌سازی اثرات هم‌تولیدی آواها بوده، این اثرات را جبران کرده و به این ترتیب به دقت بازشناسی بالاتری می‌انجامد. در مدل مذکور ابتدا یک طبقه‌بندی یا بازشناسی زمخت اولیه از فریم‌های ورودی صورت می‌گیرد، سپس در یک حرکت از بالا به پایین این تفسیرها با استفاده از یک مدل معکوس پردازش شده، با ورودی‌های سامانه مقایسه شده و در ادامه تفسیرها یا بازشناسی‌های دقیق‌تری از ورودی‌ها حاصل می‌گردد. این دسته از مدل‌ها در بهترین حالت، ۳۷٪ در دقت بازشناسی روی دادگان آزمون نسبت به مدل جلوسوی اولیه که این اثرات در آن لحاظ نشده، افزایش ایجاد کردند.

در شبکه‌ای که بر مبنای [۹] طراحی شده است (شکل ۶)، همانند نئوکورتکس در ابتدا طی یک مرحله پردازش از پایین به بالا یک تفسیر زمخت از کل ورودی‌ها صورت می‌گیرد، سپس در مرحله بعد نتایج این بازشناسی اولیه در یک مرحله پردازش از بالا به پایین، ضمن عبور از یک شبکه عصبی معکوس پردازش می‌شوند و ضمن عبور از یک شبکه عصبی جلوسو دوباره بازشناسی می‌شوند و به این ترتیب نتایج بازشناسی دقیق‌تری حاصل می‌شوند.

در این طرح ابتدا یک مدل اولیه جلوسو وجود دارد که عمل جداسازی اطلاعات فرد از حالت را انجام می‌دهد. این مدل در واقع بازشناسی زمخت اولیه از ورودی‌ها را انجام می‌دهد. در مرحله بعد یک مدل ثانویه وجود دارد که بر اساس نتایج به‌دست آمده از مدل اولیه زمینه‌ای که هر بردار ورودی در آن قرار دارد را شناسایی می‌کند. طی یک مرحله پردازش از بالا به پایین این زمینه را پردازش کرده و مدل اول را با این زمینه تطبیق می‌دهد و دوباره ورودی‌ها را با دقت بیشتری بازشناسی می‌کند. در واقع این مدل سعی می‌کند تا با استفاده از روش تطبیق مدل، بردار ورودی دارای تنوعات را با دقت بیشتری بازشناسی کند، چرا که در اثر تنوعات، ابرصفحات تصمیم‌گیرنده اشخاص یا حالات که نواحی تصمیم را شکل



شکل ۸: چند نمونه از تصاویر افراد پایگاه داده AUT در حالت نگاه به چپ و راست به همراه زاویه حالت متناظر.



شکل ۷: تصاویر مجازی شخص هشتم از پایگاه داده ZD توسط مدل پیشنهادی.

جدول ۱: نتایج بازشناسی شبکه طبقه‌بندی کننده با تصاویر مجازی حاصل از پویاکردن مدل‌های استخراج مؤلفه‌ها در مقایسه با سایر مدل‌های شبکه عصبی.

روش به کار رفته	تعداد تصاویر تعلیم	تعداد تصاویر تست	درصد صحت دادگان تعلیم	درصد صحت دادگان تست	تعداد تصاویر تست صحیح
تعلیم شبکه طبقه‌بندی کننده فقط با تصاویر نرمال افراد [۶]	۱۰	۱۱۰	۱۰۰٪	۷۰٫۹٪	۷۸
تولید تصاویر مجازی به روش خوشه‌بندی بدون سرپرستی اطلاعات افراد [۷]	۱۳۰	۱۱۰	۱۰۰٪	۸۲٫۷۳٪	۹۱
تولید تصاویر مجازی به روش خوشه‌بندی بدون سرپرستی اطلاعات افراد و حالات [۶]	۱۳۰	۱۱۰	۱۰۰٪	۸۳٫۶۳٪	۹۲
تولید تصاویر مجازی با استفاده از مدل دوسویه مبتنی بر نتوکورتکس	۱۳۰	۱۱۰	۱۰۰٪	۸۵٫۴۵٪	۹۴

تصویر ورودی را تعلیم می‌دهیم که این کار از طریق بیشینه‌کردن احتمال قطعات مرتبط انجام می‌شود. این احتمال برای قطعه تصویر تعلیم x_{np} p امین قطعه از n امین تصویر تعلیم (به صورت زیر تعریف می‌شود)

$$z_{npl}(\Phi) = \Pr(x_p, \hat{l}_n | \beta, w_{p\bullet}) = \frac{\exp(w_{pl}^T \Phi_n)}{\sum_{l=1}^L \exp(w_{pl}^T \Phi_n)} \quad (5)$$

که در آن \hat{l}_n محلی از کتابخانه است که بیش از همه با قطعه تعلیمی x_{np} انطباق دارد و به صورت زیر تعریف می‌شود

$$\hat{l}_n = \arg \max_l G_{x_{np}} [L_l, \sigma^T I] \quad (6)$$

سپس از روش کدینگ l از L برای بازنمایی محل انتخاب شده از کتابخانه (\hat{l}_n) برای هر قطعه از تصاویر تعلیم استفاده شده و در بردار هدف t_{np} ذخیره می‌شود. بنابراین بردار هدف متناظر با p امین قطعه از n امین نمونه تعلیم با بردار ویژگی Φ_n است (بردار باینری که تمامی عناصر آن به جز عنصر متناظر \hat{l}_n مساوی صفر است). با در نظر گرفتن کلیه دادگان تعلیم می‌توان (۵) را به صورت زیر بازنویسی کرد

$$\Pr(T_{\bullet p} | w_{p1}, \dots, w_{pL}) = \prod_{n=1}^N \prod_{l=1}^L \Pr(x_{np}, \hat{l}_n | \beta, w_{p\bullet})^{t_{npl}} = \prod_{n=1}^N \prod_{l=1}^L z_{npl}^{t_{npl}} \quad (7)$$

که در آن $T_{\bullet p} = z_{pl}(\Phi_n)$ و $z_{npl} = z_{pl}(\Phi_n)$ ماتریس $N \times L$ از بردارهای هدف (نزدیک‌ترین محل‌های کتابخانه) متناظر p امین قطعه و تمامی تصاویر تعلیم با المان‌های t_{npl} است (شکل ۹). اگر از (۷) لگاریتم منفی بگیریم، داریم

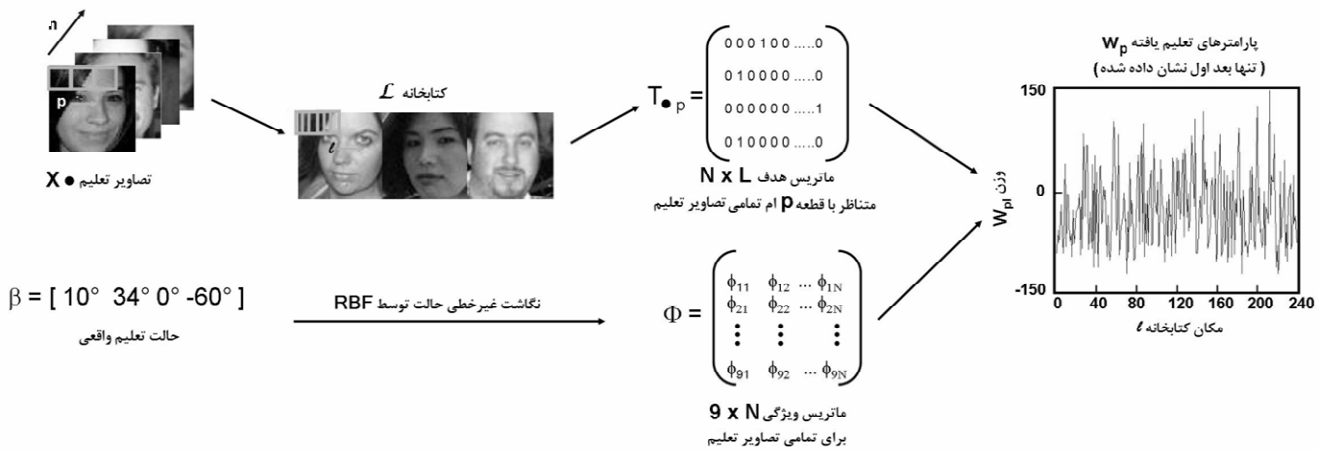
$$E(w_{p1}, \dots, w_{pL}) = -\ln \Pr(T_{\bullet p} | w_{p1}, \dots, w_{pL}) = -\sum_{n=1}^N \sum_{l=1}^L t_{npl} \ln z_{npl} \quad (8)$$

پایگاه داده ZD نمایش داده شده است. از این تصاویر مجازی جهت تعلیم شبکه عصبی طبقه‌بندی کننده استفاده می‌شود. به منظور تعلیم این شبکه با استفاده از الگوریتم پس‌انتشار خطا، چهره‌های نرمال واقعی و ۱۲ چهره مجازی تولید شده برای هر شخص به عنوان دادگان تعلیم و ۱۱ چهره واقعی هر شخص به عنوان داده تست در نظر گرفته شده است. نتایج به دست آمده از این آزمایش در جدول ۱ آمده است. همان‌گونه که مشاهده می‌شود، درصد صحت بازشناسی نسبت به شبکه جلوسوی خوشه‌بندی اطلاعات افراد و حالات، ۱/۸۲٪ رشد کرده است که نشان‌دهنده بهبود جداسازی مؤلفه‌های فرد و حالت است.

۴-۱ ارزیابی کمی مدل پیشنهادی

در پایگاه داده AUT حالات نگاه به راست و چپ برای افراد مختلف دارای زاویه یکسان نمی‌باشند و همین مسئله موجب می‌شود که شبکه با حالت چند به یک و تناقض مواجه شود، چرا که مثلاً برای حالت نگاه به راست باید کد واحدی را اختصاص دهد (شکل ۸). در اینجا نشان می‌دهیم که این مشکل تا حدودی توسط شبکه بازگشتی رفع می‌شود، زیرا پس از جداسازی، مؤلفه‌ها دوباره توسط شبکه معکوس به ورودی داده می‌شوند و اصلاح می‌گردند. برای این منظور از روش تخمین زاویه حالت چهره بر مبنای بازنمایی چهره با شبکه‌ای از قطعات غیر هم‌پوشان استفاده شده است [۱۶]. در واقع در این روش از یک بازنمایی عمومی استفاده شده که به مکان ویژگی‌های چهره یا تطبیق یک مدل به چهره وابسته نیست. در این رویکرد تصویر تست به شبکه‌ای از قطعات شکسته می‌شود که هر یک به صورت جداگانه شامل اطلاعات مستقلی درباره حالت واقعی چهره است. هسته الگوریتم مورد استفاده یک کتابخانه از پیش تعریف شده است و از ارتباط بین قطعات چهره تست و قطعات موجود در کتابخانه برای تخمین حالت چهره استفاده می‌شود.

برای تعلیم این مدل، یک بردار وزن جداگانه $w_{p\bullet}$ برای هر قطعه از



شکل ۹: تعلیم مدل با N تصویر تعلیم انجام می‌شود که هر کدام به صورت شبکه‌ای از قطعات غیر هم‌پوشان بیان شده‌اند. کتابخانه L را می‌توان به صورت مجموعه‌ای از قطعات L_i در نظر گرفت که منظور از ℓ ، L محل ممکن است. از طرح کدینگ یک از L استفاده می‌کنیم تا نزدیک‌ترین قطعات کتابخانه را در ماتریس هدف T_p ذخیره کنیم. مقادیر حالت واقعی در بردار β ذخیره می‌شوند که برداری $N \times 1$ است [۱۶].

دوسویه اطلاعات فرد و حالت ارائه شد. به کمک شبکه عصبی تحلیل‌گر به جداسازی اطلاعات فرد از حالت و تخمین مانیفولدهای زیرفضاهای مربوطه پرداختیم و نشان دادیم که با افزودن شبکه معکوس، مدل دارای خاصیت تطبیقی خواهد بود و در نتیجه مؤلفه‌های استخراج‌شده به مؤلفه‌های واقعی نزدیک‌تر می‌شوند. پس از تولید تصاویر مجازی به کمک شبکه تحلیل‌گر از این تصاویر جهت افزایش دادگان تعلیم در شبکه عصبی طبقه‌بندی کننده استفاده شد. در توضیح بهبود کیفیت تصاویر مجازی حاصل و افزایش درصد صحت بازشناسی شبکه عصبی طبقه‌بندی کننده، می‌توان گفت همان‌گونه که مغز انسان با شناختن افراد، حالات آنها را بهتر تشخیص می‌دهد، در اینجا نیز از طریق شبکه معکوس و دادن اطلاعات بخش افراد به بخش حالات، زیرفضاهای اطلاعات افراد و حالات با یکدیگر تعامل خواهند داشت و این اثر متقابل، در بازیابی و تداعی اطلاعات و همچنین افزایش قدرت تعمیم شبکه مؤثر است.

در ادامه با استفاده از تخمین زاویه حالت چهره بر مبنای بازیابی چهره با شبکه‌ای از قطعات غیر هم‌پوشان به بررسی زاویه حالت تصاویر مجازی حاصل پرداختیم و نشان دادیم که نرمالیزاسیون حالت و یا همسوسازی انجام گرفته است. زیرا با به کارگیری شبکه بازگشتی، مؤلفه‌ها پس از جداسازی، دوباره توسط شبکه معکوس به ورودی داده شده و اصلاح می‌شوند.

باید توجه داشت که تعلیم مناسب فضاهای چهره اشخاص و حالات به کمک شبکه نیازمند به کارگیری مدل‌هایی از شبکه‌های عصبی با اتصالات بازگشتی در محل‌های مناسب است که همانند نتوکورتکس عمل بازشناسی طی مراحل متعددی از بازشناسی جلوسو و برگشت و دورزدن در شبکه صورت پذیرد. در تحقیقات آینده می‌توان اثر افزودن اتصالات بازگشتی و دادن دوباره مؤلفه‌های بخش افراد و حالات به همان بخش و تعامل درون مانیفولدها را بر جداسازی اطلاعات افراد و حالات مورد بررسی قرار داد. اگر تخمین دقیق‌تری از این مانیفولدها حاصل شود، می‌توان با داشتن مانیفولدهای تغییر حالت در چهره و به دست آوردن محل تلاقی آنها با مانیفولدهای تغییر شخص، حالات مجازی شخص را به صورت مناسب‌تری تولید کرد. همچنین باید مشخص کرد که مؤلفه‌های استخراج‌شده افراد و حالات متعلق به کدام دسته از مؤلفه‌های اساسی و مستقل می‌باشند.

جدول ۲: زاویه حالت تصاویر مجازی و تصاویر واقعی برای حالت نگاه به راست.

شماره تصویر در پایگاه داده ZD	زاویه حالت تصویر مجازی	زاویه حالت تصویر واقعی در پایگاه داده
۱	۳۴	۵۲
۴	۲۶	۴۵
۸	۳۲	۳۹
۱۰	۲۸	۳۰

در مرحله آخر برای تعیین پارامترهای W_{pl} ، (λ) کمینه می‌شود که این کار با گرفتن گرادیان از (λ) نسبت به یکی از بردارهای پارامتر W_{pl} قابل انجام است

$$\nabla_{w_{pl}} E(w_{p1}, \dots, w_{pL}) = \sum_{n=1}^N (z_{npl} - t_{npl}) \Phi_n \quad (9)$$

از آنجایی که (λ) به تمامی پارامترهای $w_{p1} \dots w_{pL}$ بستگی دارد، تمامی گرادیان‌های $\nabla_{w_{pl}}$ با هم ترکیب می‌شوند تا این عمل معادل مشتق‌گیری باشد و این کار برای هر قطعه تکرار می‌شود.

جهت تعلیم مدل از پایگاه داده تصاویر جمع‌آوری شده توسط گروه بینایی ماشین دانشگاه کالج لندن استفاده شده است [۱۶]. این پایگاه داده دارای تنوعات وسیعی به لحاظ نورپردازی، مقیاس، حالت چهره و غیره است. از ۱۰۹۰۰ تصویر تعلیم و ۱۰۰۰ تصویر تست استفاده شده و کتابخانه نیز شامل ۲۴۰ تصویر است.

با بررسی‌های انجام‌شده و محاسبه زاویه تصاویر سنتز شده توسط شبکه، مشاهده می‌شود که زاویه این تصاویر در بازه ۲۵ تا ۳۵ درجه است. البته باید در نظر داشت که به دلیل رزولوشن پایین تصاویر سنتز شده در برخی حالات، امکان وجود خطا تا حدودی وجود دارد، ولی سعی شده که روش انتخاب‌شده برای تخمین حالت نسبت به تغییرات شرایط مقاوم باشد. در جدول ۲ زاویه حالت تصاویر مجازی و تصاویر واقعی برای حالت نگاه به راست آمده است. همان‌گونه که مشاهده می‌شود، در تصاویر مجازی حاصل تقریباً به نوعی نرمالیزاسیون حالت صورت گرفته است.

۵- جمع‌بندی و نتیجه‌گیری

در این مقاله با هدف بهبود بازشناسی چهره با یک تصویر از هر فرد، با الهام از مدل نتوکورتکس یک ساختار شبکه عصبی جهت پردازش

- [10] I. Nejadgholi and S. Seyyedsalehi, "Nonlinear normalization of input patterns to speaker variability in speech recognition neural networks," *Neural Computing & Applications*, vol. 18, no. 1, pp. 45-55, Jan. 2009.
- [11] M. M. Mesulam, "From sensation to cognition," *Brain*, vol. 121, no. 6, pp. 1013-1052, Jun. 1998.
- [12] E. Koerner and G. Matsumoto, "Cortical architecture and self-referential control for brain-like computation, a new approach to understanding how the brain organizes computation," *IEEE Eng. in Medicine and Biology Magazine*, vol. 21, no. 5, pp. 121-133, Sep./Oct. 2002.
- [13] E. Koerner, H. Tsujino, and T. Masutani, "A cortical-type modular neural network for hypothetical reasoning," *Neural Networks*, vol. 10, no. 5, pp. 791-814, Jul. 1997.
- [14] E. Koerner and et al., "A model of computation in neocortical architecture," *Neural Networks*, vol. 12, no. 12, pp. 989-1005, Oct. 1999.
- [15] J. Ghosen and Y. Bengio, "Bias learning, knowledge sharing," *IEEE Trans. on Neural Networks*, vol. 14, no. 4, pp. 748-765, Jul. 2003.
- [16] J. Aghajanian and S. J. D. Prince, "Face pose estimation in uncontrolled environments," in *Proc. British Machine Vision Conf.*, 2009.

فاطمه عبدالعلی مدرک کارشناسی خود را در مهندسی پزشکی - بیوالکترونیک از دانشگاه صنعتی امیرکبیر در سال ۱۳۸۷ و مدرک کارشناسی ارشد را در همان رشته از دانشگاه صنعتی امیرکبیر در سال ۱۳۸۹ دریافت کرده است. وی اکنون مشغول به تحصیل در دوره دکتری مهندسی پزشکی - بیوالکترونیک در دانشگاه تهران می‌باشد. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: پردازش سیگنال با بهره‌گیری از هوش مصنوعی، مدل‌سازی عملکرد مغز و شبکه‌های عصبی مصنوعی.

سیدعلی سیدصالحی مدرک کارشناسی خود را در مهندسی برق از دانشگاه صنعتی شریف در سال ۱۳۶۱، کارشناسی ارشد را در مهندسی برق از دانشگاه صنعتی امیرکبیر در سال ۱۳۶۷ و دکتری خود را در مهندسی برق - بیوالکترونیک از دانشگاه تربیت مدرس در سال ۱۳۷۴ دریافت نموده است. وی در حال حاضر دانشیار دانشکده مهندسی پزشکی دانشگاه صنعتی امیرکبیر می‌باشد. زمینه‌های پژوهشی مورد علاقه ایشان پردازش و بازشناسی گفتار، شبکه‌های عصبی مصنوعی و زیستی، مدل‌سازی عملکرد مغز و پردازش خطی و غیر خطی سیگنال می‌باشد.

۶- سپاس‌گزاری

بدین وسیله از زحمات آقای دکتر کریم فائز و متخصصین آزمایشگاه پردازش تصویر و شناسایی الگوی دانشکده برق دانشگاه صنعتی امیرکبیر که دادگان AUT را در اختیار ما قرار دادند و همچنین از خانم مهندس زمانی که دادگان ZD را طی پروژه کارشناسی خود در دانشکده مهندسی پزشکی جمع‌آوری نمودند، سپاس‌گزاری می‌شود.

مراجع

- [1] X. Zhang and Y. Gao, "Face recognition across pose: a review," *Pattern Recognition*, vol. 42, no. 11, pp. 2876-2896, Nov. 2009.
- [2] T. Stonham, "Practical face recognition and verification with WISARD," *Aspects of Face Processing*, pp. 426-441, 1984.
- [3] S. H. Lin, S. Y. Kung, and L. J. Lin, "Face recognition/detection by probabilistic decision-based neural network," *IEEE Trans. Neural Network*, vol. 8, no. 1, pp. 114-132, Jan. 1997.
- [4] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: a convolutional neural-network approach," *IEEE Trans. Neural Network*, vol. 8, no. 1, pp. 98-113, Jan. 1997.
- [5] Y. W. Wong, K. P. Seng, and L. M. Ang, "Radial basis function neural network with incremental learning for face recognition," *IEEE Trans. Syst. Man Cybern. B*, vol. 41, no. 4, pp. 940-949, Aug. 2011.
- [6] F. Abdolali and S. Seyyedsalehi, "Face recognition from a single image per person using deep architecture neural networks," in *Proc. 3rd Int. Conf. on Computer and Electrical Engineering, Chengdu*, vol. 1, pp. 70-73, Nov. 2010.
- [7] ن. داداشی، *بازشناسی چهره با استفاده از یک تصویر از هر فرد*، پایان‌نامه کارشناسی ارشد، دانشگاه صنعتی امیرکبیر، دانشکده مهندسی پزشکی، دی ۱۳۸۷.
- [8] ع. سیدصالحی، افزایش کارایی بازشناخت الگوی شبکه‌های عصبی جلوسو از طریق توسعه روش‌هایی برای دوسویه‌کردن عملکرد آنها، گزارش طرح پژوهشی، دانشگاه صنعتی امیرکبیر، دانشکده مهندسی پزشکی، ۱۳۸۳.
- [9] A. Leila and S. Seyyedsalehi, "Modeling phones coarticulation effects in a neural network based speech recognition system," in *Proc. INTERSPEECH-2004*, vol. 1, pp. 1657-1660, Oct. 2004.