بهبود مدل تفکیک‌کنندهٔ منیفلدهای غیرخطی به‌منظور بازشناسی چهره با یک تصویر از هر فرد

سیّده زهره سیّدصالحی و سیّد علی سیّدصالحی
دانشکدهٔ مهندسی پزشکی، دانشگاه صنعتی امیرکبیر، تهران، ایران

چکیده

یادگیری منیفلد یکی از روش‌های کاهش بعد مطرح به‌منظور استخراج ساختار غیرخطی داده با ابعاد بالاست. تاکنون روش‌های زیادی به این منظور ارائه شده‌اند. در تمام این روش‌ها یک منیفلد به‌عنوان منیفلد جاسازی‌شده در داده استخراج می‌شود. درحالی‌که در خیلی از مسائل مربوط به دنیای واقعی یک منیفلد به‌تنهایی بیان‌گر ساختار داده نیست. در این راستا بر مبنای تحقیقات قبلی، یک روش کاهش بعد غیرخطی مبتنی بر شبکه‌های عصبی عمیق ارائه شده است که قادر به استخراج توأم منیفلدهای جاسازی‌شده در داده است. در مدل شبکهٔ عصبی تفکیک‌کنندهٔ منیفلدهای غیرخطی، برخلاف روش معمول استخراج منیفلد با شبکه‌های عصبی که به‌صورت بدون سرپرستی صورت می‌گیرد، از برچسب داده در جهت شکل‌گیری منیفلدها به‌صورت غیرمستقیم استفاده می‌شود. با توجه به ساختار عمیق این مدل نشان داده شده است که با بهره‌گیری از روش‌های پیش‌تعلیم می‌توان به‌طور معناداری عملکرد آن‌را بهبود بخشید؛ همچنین در راستای استخراج بهتر منیفلدها و حفظ تمایز درون‌منیفلدی برای طبقات مختلف، توابع معیار آن بهبود داده شده است. این مدل برای استخراج منیفلدهای حالت‌های احساسی و هویت افراد از دادگان چهره +CK، مورد استفاده قرار گرفته است. با بهره‌گیری از پیش‌تعلیم لایه‌به‌لایه و بهبود توابع معیار، نرخ بازشناسی حالت برای تصاویر مجازی از ٪۲۴/۲۹ به ٪۷۵/۰۷ و درصد صحت بازشناسی هویت با یک تصویر از هر فرد با غنی‌سازی دادگان تعلیم طبقه‌بند KNN توسط این تصاویر مجازی، از ٪۹۰/۶۳ به ٪۹۷/۰۷ نسبت به مدل اولیه بهبود داشته است.

واژه‌گان کلیدی: شبکهٔ عصبی، یادگیری منیفلد، تمایز درون‌منیفلدی، الگوهای مجازی، ساختار عمیق، تفکیک منیفلدها.

۱- مقدمه

در دو دههٔ اخیر شناسایی چهره مورد توجه تحقیقات وسیعی از بینایی رایانه و شناسایی الگو بوده است. یکی از کاربردهای وسیع بازشناسی چهره در زمینه تأیید هویت و مسألهٔ امنیت است. در کنترل اماکن با جمعیت زیاد مانند فرودگاه‌ها، این روش نسبت به سایر روش‌های نظارتی کاراتر است. چرا که برخلاف برخی از آنها مانند اثر انگشت یا بازشناسی عنبیه به همکاری سوژه نیاز ندارد. در این روش از چهرهٔ افراد عکس‌های مختلفی گرفته شده و دستگاه باید توانایی شناسایی این افراد را در زمان‌ها و ژست‌های متفاوت، زوایای تابش نور مختلف و... داشته باشد (داداشی، ۱۳۸۷).

بیشتر روش‌های بازشناسی چهره نیاز به تعدادی عکس از چهرهٔ هر فرد برای آموزش دارند درحالی‌که در بسیاری از موارد، چندین عکس از یک فرد در دسترس نیست. برای بهبود این مشکل پژوهش‌گران، حل مسألهٔ بازشناسی تنها با یک تصویر از چهره را مطرح کردند (Tan, Chen et al. 2006, Wang, Li et al. 2013).

در بازشناسی چهره با استفاده از یک تصویر از هر فرد، از هر شخص فقط یک تصویر خنثی برای تعلیم موجود بوده و تصاویر آزمون در حالت‌ها یا وضعیت‌های مختلف هستند. بنابراین به طبقه‌بند فقط یک تصویر خنثی تعلیم داده می‌شود. مشکل اصلی در اینجا عدم توانایی یک تصویر برای ایجاد تعمیم لازم برای بازشناسی است. برای حل این مسأله در سال‌های اخیر روش‌های زیادی از جمله تولید

نمونه‌های مجازی[1] (داداشی و همکاران، ۱۳۹۰؛ نژادقلی، ۱۳۹۱, Hsieh, Lai et al. 2010, Abdolali and Seyyedsalehi 2012, Mohammadzade and Hatzinakos 2013, Zhu, Tang et al. 2013)، محلی‌سازی تک تصویر تعلیمی[2] (Chen, Liu et al. 2004) و تطابق احتمالاتی[3] (Martínez 2002, Martınez 2003) پیشنهاد شده‌اند. رویکردی که در مقالهٔ حاضر برای بهبود بازشناسی چهره با یک تصویر از هر فرد، مد نظر قرار گرفته است، روش تولید تصاویر مجازی است.

در (Mohammadzade and Hatzinakos 2013) با استفاده از تحلیل تفکیک خطی، زیرفضای هر یک از حالت‌های احساسی ساخته و با نگاشت تصویر خنثی آزمون در هر یک از این زیرفضاها حالت‌های مجازی مختلف برای این تصویر، تولید شده است. (Kan, Shan et al. 2013) نیز روش تجزیه و تحلیل تفکیک‌کنندهٔ تطبیقی[4] را برای تولید تصاویر مجازی پیشنهاد داده است. این روش با یافتن نمونه‌های مشابه به تصویر آزمون از دادگان تعلیم نمونه‌های مجازی را تخمین می‌زند.

در (داداشی، ۱۳۸۷) از رویکرد جداسازی اطلاعات هویت از حالت و تخمین منیفلدهای مربوطه به کمک شبکه‌های عصبی جلوسوی چندلایه به‌منظور بازشناسی چهره با یک تصویر از هر فرد استفاده شده است. در این روش با به‌کارگیری منیفلدهای تخمین‌زده‌شده و تولید تصاویر مجازی از طریق این منیفلدها، به افزایش دادهٔ تعلیم سامانهٔ بازشناس فرد پرداخته شده است. منیفلد افراد در این ساختار با یک روش خوشه‌بندی تخمین زده می‌شود؛ اما منیفلدی که این مدل برای حالت ارائه می‌کند، در عمل برچسب‌های حالتی هستند که توسط کاربر تعریف شده است و نمی‌توان آن را به‌عنوان منیفلد ذاتی حالت در نظر گرفت. در ادامهٔ این تحقیق در (عبدالعلی، ۱۳۸۹؛ داداشی و همکاران، ۱۳۹۰) مدل توسعه داده شده و برای استخراج هر دو منیفلد هویت و حالت از روش خوشه‌بندی به‌کار رفته در (داداشی، ۱۳۸۷) استفاده شده که در نتیجهٔ آن، نرخ بازشناسی در سامانهٔ بازشناس فرد بهبود یافته است.

در مقالهٔ حاضر به شرح کامل و توسعه مدل ارائه‌شده در (داداشی و همکاران، ۱۳۹۰) با عنوان شبکهٔ عصبی تفکیک‌کنندهٔ منیفلدهای غیرخطی[5] (NMSNN) پرداخته و نشان داده شده است که با استفاده از روش پیش‌تعلیم لایه‌به‌لایه (سیّدصالحی و سیّدصالحی، ۱۳۹۲)، قابلیت این مدل به‌طور معناداری افزایش می‌یابد. چراکه با توجه به ساختار عمیق مدل (Bengio, 2012)، بدون بهره‌گیری از روش‌های پیش‌تعلیم، کمینه‌های محلی مانع از یادگیری مطلوب مؤلفه‌ها در عمق می‌شوند؛ همچنین به‌منظور افزایش تمایز درون‌منیفلدی بین طبقات مختلف، بخش دیگری به تابع معیار یادگیری منیفلدها افزوده شده که در دو گام نیز بهبود داده شده است. وجود این بخش در مواردی‌که شباهت‌های برون‌طبقه‌ای به درون‌طبقه‌ای در فضای یک منیفلد بیشتر باشد ضروریست.

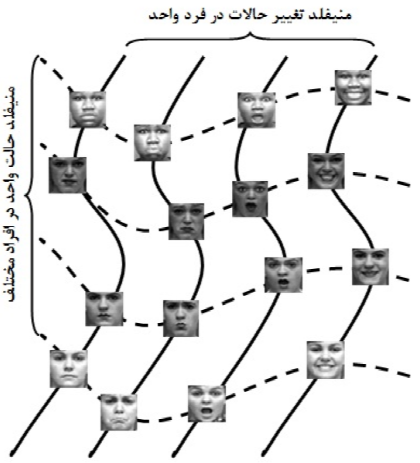
NMSNN با جداسازی زیرمنیفلدهای غیرخطی اطلاعات مربوط به هویت افراد از اطلاعات مربوط به حالت آنها، امکان تعمیم‌دهی یک حالت چهره از یک فرد به سایر افراد و نیز تولید سایر حالت‌های یک فرد از چهره آن را فراهم می‌کند. از دیدگاه دیگر، استخراج مؤلّفه‌های مشترک قابل تعمیم از داده‌ها، زمینهٔ تجزیه و تحلیل غیرخطی اطلاعات نهفته در دادگان را ایجاد می‌کند. این مؤلّفه‌های مشترک به‌عبارت دیگر ویژگی‌های غیرخطی اساسی قابل تعمیم درون‌داده‌ها می‌باشند که شواهد نشان می‌دهد مغز ما در تجزیه و تحلیل اطلاعات ورودی، از آنها بسیار استفاده می‌کند (Bengio, 2009). با استخراج مؤلّفه‌های غیرخطی واقعی درون داده‌ها، امکان شبیه‌سازی تصور و تخیل توسط ذهن انسان، در شبکه‌های عصبی مصنوعی فراهم می‌شود. به‌نظر می‌رسد این مؤلّفه‌های مشترک قابل تعمیم، زیربنای عملکرد مغز انسان در تفکر و خلاقیت نیز هستند.

برای روشن‌تر شدن مسأله داده‌هایی را که متشکل از حالات مشترکی بین چهره افراد مختلف هستند، در نظر بگیرید؛ در روش‌های معمول، هر تصویر به‌عنوان یک الگوی مجزای یادگیری در نظر گرفته می‌شود. این درحالیست که بین تصاویر، الگوهای مشترکی وجود دارند که می‌توانند به‌عنوان دانش اولیه در ساختار مدل به‌کار گرفته شوند. به‌عبارتی می‌توان به‌جای تعلیم هر الگوی تصویر به‌صورت مجزا و استخراج یک منیفلد کلی برای داده، دو زیرمنیفلد استخراج کرد، به‌گونه‌ای‌که هریک از این زیرمنیفلدها حاوی اطلاعات مربوط به تغییرات خاصی در الگوها باشند (Seyyedsalehi and Seyyedsalehi, 2014).

به‌طور مثال برای داده (شکل ۱-۱) به‌جای استخراج یک منیفلد کلی، با توجه به اینکه الگوهای هریک از حالت‌های احساسی برای افراد مختلف ثابت هستند و به‌طور

[1] Synthesizing Virtual Samples
[2] Localizing the Single Training Image
[3] Probabilistic Matching
[4] Adaptive Discriminant Analysis
[5] Nonlinear Manifold Separator Neural Network

برعکس برای الگوی فرد در حالت‌های احساسی مختلف، می‌توان دو زیرمنیفلد حالات و اشخاص را در نظر گرفت. با به اشتراک گذاشتن دانش بین این دو زیرمنیفلد، تصاویر افراد مختلف با حالات متنوع قابل بازسازی است. یک کاربرد مهم این روش فراهم‌شدن امکان تولید تصاویر مجازی است؛ یعنی چهره‌های افراد را در حالت‌هایی که در دادگان وجود ندارد، به‌طور مصنوعی می‌توان تولید کرد. این کار تا حدودی مشابه حالت تصور و تخیل و خلاقیت در انسان است. این در حالیست که روش‌های رایج استخراج‌کننده تک‌منیفلد فاقد این توانمندی هستند؛ اما NMSNN با استخراج هم‌زمان منیفلدهای حالت و هویت، امکان به اشتراک گذاشته‌شدن مؤلفه‌های حالت را جهت بازسازی تصاویر مجازی و غنی‌سازی دادگان تعلیم برای بازشناسی هویت با یک تصویر از هر فرد فراهم می‌کند.

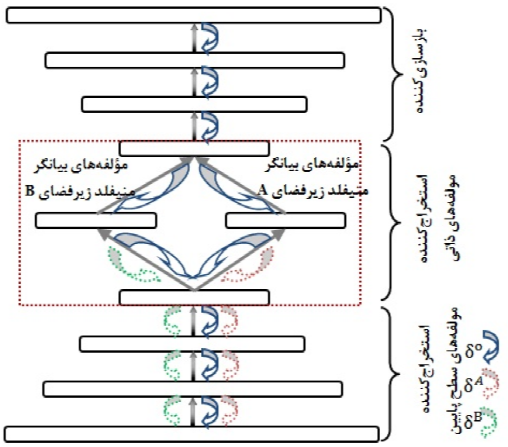


(شکل ۱-۱): نمایش شماتیک از منیفلدهای تغییر حالات و افراد.

در ادامه به شرح ساختار شبکهٔ عصبی تفکیک‌کنندهٔ منیفلدهای غیرخطی پرداخته می‌شود؛ سپس در بخش سوم توابع معیار برای بهبود تفکیک درون‌منیفلدی معرفی می‌شوند. در بخش چهارم به روش پیش‌تعلیم به‌کار رفته اشاره می‌شود. بخش پنجم نیز سامانهٔ بازشناس فرد را معرفی می‌کند. در بخش ششم هم به شرح پیاده‌سازی‌ها و نتایج آنها پرداخته و درنهایت جمع‌بندی و نتیجه‌گیری در بخش هفتم آورده شده است.

۲- شبکهٔ عصبی تفکیک‌کننده منیفلدهای غیرخطی (NMSNN)

NMSNN یک شبکهٔ عصبی چندلایه با ساختار عمیق است که در (شکل ۲-۱) ساختار آن معرفی شده است. ساختار اولیهٔ این مدل به‌عنوان مدل مرجع، NMSNN0 نامیده شده است. این مدل برای استخراج توأم منیفلدهای دادگانی که چندین منیفلد دارند، کاراست. ساختار آن مشابه یک شبکهٔ عصبی گلوگاه[6] (BNN) است با این تفاوت که نورون‌های لایهٔ گلوگاه آن به دو بخش تقسیم شده‌اند که هر بخش آن، وظیفه یادگیری یکی از منیفلدهای موردنظر را به عهده دارند؛ البته این مدل برای استخراج توأم بیش از دو منیفلد هم قابل تعمیم است که وابسته به داده‌ها می‌توان آن را توسعه داد. اما در این مقاله ساختاری برای استخراج توأم دو منیفلد ارائه شده است.



(شکل ۲-۱): ساختار مدل NMSNN0

همان‌طور که در (شکل ۲-۱) نشان داده شده است، مدل NMSNN0 را به سه بخش می‌توان تقسیم کرد، بخش‌های استخراج‌کنندهٔ مؤلفه‌های سطح پایین، استخراج‌کنندهٔ مؤلفه‌های ذاتی و بازسازی‌کننده. بخش استخراج‌کنندهٔ مؤلفه‌های سطح پایین در طی چندین مرحله مؤلّفه‌های اساسی غیرخطی داده را استخراج می‌کند. باید توجه داشت که این مؤلفه‌ها بیان‌گر منیفلد کلی داده هستند که سایر روش‌های غیرخطی استخراج منیفلد قادر به استخراج آن هستند؛ اما آنچه که قابلیت اصلی مدل NMSNN0 است استخراج توأم چند منیفلد از درون داده

[6] Bottleneck Neural Network

است که در بسیاری از کاربردهای واقعی، تنها یک منیفلد کلی نمی‌تواند بیان‌گر اطلاعات موجود باشد؛ بلکه تعدادی منیفلد با به اشتراک گذاشتن دانش مابین هم، قادر به بیان اطلاعات موجود خواهند بود (Seow, 2006)؛ لذا در ادامه در بخش استخراج‌کنندهٔ مؤلفه‌های ذاتی سعی می‌شود که از درون این منیفلد کلی، زیرمنیفلدهای جاسازی‌شده در داده استخراج شوند که بر حسب نوع داده، زیرمنیفلدهای مختلفی می‌توان تعریف کرد. در (شکل ۲-۱) دادهٔ ورودی به‌طور کلی ترکیب غیرخطی دو منیفلد A و B فرض شده است که مدل به استخراج آن دو می‌پردازد. در این بخش لایهٔ میانی به دو بخش تقسیم شده است که هر بخش آن، وظیفهٔ یادگیری مؤلفه‌های مربوط به یکی از زیرفضاهای A و B را به‌عهده دارد. در بخش سوم نیز با ترکیب غیرخطی منیفلدهای A و B داده الگوی ورودی بازسازی می‌شود.

در طراحی این ساختار دو نکتهٔ اصلی مد نظر قرار داده شده است. ابتدا اینکه طراحی ساختار به‌نحوی صورت گرفته است که امکان به اشتراک گذاشته‌شدن دانش[7] در آن فراهم گردد. این به اشتراک گذاشته‌شدن دانش مربوط به مؤلفه‌ها در مدل‌های مورد یادگیری، از جهات مختلفی مورد اهمیت است. از یک سو مؤلفه‌ها و مفاهیم مورد یادگیری می‌توانند از انواع متنوعی از داده‌ها تعلیم ببینند و به این ترتیب نیاز به داشتن دادگان بسیار بزرگ از یک نوع خاص برطرف می‌شود. از سوی دیگر زمینهٔ خلاقیت و تولید تنوعات جدید از الگوی مورد تعلیم در مدل فراهم می‌شود؛ بدون آنکه این تنوعات جدید از الگوها را مدل در قبل دیده باشد. به این ترتیب به اشتراک گذاشته‌شدن دانش بین منیفلدها منجر به یادگیری بهتر و تعمیم‌دهی دقیق‌تر به‌خصوص در مواقعی‌که داده آموزشی محدود است، می‌شود. دوم این‌که با اعمال دانش اولیه مناسب در مسیر تعلیم NMSNN0، به‌شکل‌گیری منیفلدها در راستای هدف موردنظر جهت‌دهی شده است. این اعمال دانش اولیه با استفاده غیرمستقیم از اطلاعات برچسب نمونه‌ها در طی تعلیم صورت گرفته است که در ادامه شرح داده می‌شود.

برای تعلیم NMSNN0 سه تابع هزینه تعریف می‌شود که در روابط (۱) الی (۳) آورده شده‌اند. الگوریتم یادگیری باید به‌گونه‌ای تنظیم شود که هر سه تابع هزینه را هم‌راه باهم کمینه کند. با توجه به اینکه فرض شده دادهٔ محصول ترکیب غیرخطی دو منیفلد A و B است؛ لذا برای هر نمونه دو برچسب متناسب با هر یک از دو منیفلد در نظر گرفته می‌شود.

$$E^{O} = \frac{1}{2}\sum_{i=1}^{n_o}(D_i - Y_{oi})^2 \qquad (۱)$$

E^{O} در رابطه (۱) تابع معیار خطای خودانجمنی است که مجموع مربعات خطای بازسازی الگوی ورودی در لایهٔ خروجی است. در این رابطه D و Y_o به‌ترتیب خروجی مطلوب (همان ورودی) و خروجی بازسازی‌شده توسط شبکه هستند که o اندیس لایهٔ خروجی و n_o تعداد نورون‌های این لایه است. این خطا از لایهٔ خروجی پس‌انتشار می‌شود و در اصلاح تمام وزن‌های ساختار دخیل است. E^{O} مربوط به بخش بدون سرپرستی استخراج منیفلدهاست که در روش استخراج تک‌منیفلدی توسط BNN نیز به‌کار می‌رود. این خطا تضمین می‌کند که منیفلدهای A و B به‌گونه‌ای استخراج گردند که با ترکیب غیرخطی آن‌ها امکان بازسازی مجدد داده وجود داشته باشد؛ یعنی حاوی مؤلفه‌های اساسی داده باشند.

$$E^{A} = \frac{1}{2}\sum_{i=1}^{n_A}\left(M_{lai}^{A} - Y_{ci}\right)^2 \qquad (۲)$$

E^{A} در رابطهٔ (۲) تابع معیار خطای استخراج منیفلد A است. $Y_{c(1:n_A)}$ خروجی شبکه در بخش استخراج منیفلد A و M_A بیان‌گر منیفلد A و M_{la}^{A} مکانی روی این منیفلد متناظر با برچسب نمونه آموزشی ارائه شده به شبکه، مربوط به منیفلد A، یعنی la است. این خطا براساس نمونه‌هایی که در فضای منیفلد A باید بیان یکسان داشته باشند، تعیین می‌شود که مربوط به بخش باسرپرستی استخراج منیفلدهاست. خطای بالا تضمین می‌کند در این فضا فقط مؤلفه‌های تعریف‌شده برای منیفلد A تغییر کنند.

$$E^{B} = \frac{1}{2}\sum_{i=1}^{n_B}\left(M_{lbi}^{B} - Y_{c(i+n_A)}\right)^2 \qquad (۳)$$

E^{B} در رابطهٔ (۳) تابع معیار خطای استخراج منیفلد B است. $Y_{c(n_A+1:n_A+n_B)}$ خروجی شبکه در بخش استخراج منیفلد B و M_B بیان‌گر منیفلد B و M_{lb}^{B} مکانی روی این منیفلد متناظر با برچسب نمونهٔ آموزشی ارائه شده به شبکه مربوط به منیفلد B است. این خطا تضمین می‌کند که در زیرفضای منیفلد B نیز فقط مؤلفه‌های تعریف‌شده برای منیفلد B تغییر کنند. منیفلدها با روابط (۴) و (۵)

[7] Knowledge Sharing

به‌روزرسانی می‌شوند که در آنها $0 < \gamma < 1$ یک ضریب ثابت است.

$$M_{la}^{A}(t) = \gamma M_{la}^{A}(t-1) + (1-\gamma)Y_{c(1:n_A)} \quad (۴)$$

$$M_{lb}^{B}(t) = \gamma M_{lb}^{B}(t-1) + (1-\gamma)Y_{c(1+n_A:n_A+n_B)} \quad (۵)$$

در این روابط بخشی از منیفلد که متناظر با برچسب ورودی است، به‌روز می‌شود. می‌توان نشان داد که با این روش عملاً یک میانگین‌گیری برخط وزندار روی ورودی‌های هم‌برچسب که باید در زیرفضای مربوطه بیان یکسانی داشته باشند صورت می‌گیرد. بدین‌منظور در رابطه (۶) اگر $M_{la}^{A}(t-1), \dots, M_{la}^{A}(1)$ جایگذاری شوند، این رابطه را می‌توان نوشت:

$$M_{la}^{A}(t) = \gamma^{t} M_{la}^{A}(0) + \sum_{r=1}^{t} \gamma^{t-r}(1-\gamma)Y_{c(1:n_A)}(r) \quad (۶)$$

که در آن $M_{la}^{A}(0)$، شامل مقادیر تصادفی در بازه (۰ و ۱) می‌باشد. با توجه به $0 < \gamma < 1$، در رابطه (۶) از جمله اول می‌توان صرف‌نظر کرد. لذا:

$$M_{la}^{A}(t) \cong \sum_{r=1}^{t} \gamma^{t-r}(1-\gamma)Y_{c(1:n_A)}(r) \quad (۷)$$

این بدان معناست که در این میانگین‌گیری برخط، نمونه‌های اخیر که مربوط به مراحل تعلیم‌یافته‌تر شبکه هستند، از وزن بالاتری برخوردار هستند. در این مقاله، این روش، خوشه‌بندی باسرپرستی نامیده شده است. توابع هزینه (۲) و (۳) در جهت هدایت الگوهای هم‌برچسب در خروجی بخش استخراج هر منیفلد به سمت یکدیگر یعنی این میانگین‌ها عمل می‌کنند.

به‌منظور بهینه‌سازی توابع هزینه، لازم است اصلاح وزن‌های شبکه در جهت عکس گرادیان هر یک از خطاهای تحت تأثیر آن وزن، صورت بگیرد. فرض کنید که E^{c} یکی از توابع هزینه تعریف شده باشد، آنگاه سهم این خطا در تصحیح $W_{(j-1)j}(r,s)$ مطابق رابطهٔ (۸) است:

$$\Delta W_{(j-1)j}(r,s) = -\eta \frac{\partial E^{c}}{\partial W_{(j-1)j}(r,s)} \quad (۸)$$

$W_{(j-1)j}(r,s)$ وزن اتصال بین نورون‌های rام از لایه $j-1$ و sام از لایه j است. η نیز ضریب یادگیری می‌باشد. گرادیان تابع خطا E^{c} در رابطه (۹) محاسبه شده است.

$$\frac{\partial E^{c}}{\partial W_{(j-1)j}(r,s)} = \frac{\partial E^{c}}{\partial Y_{ji}} \cdot \frac{\partial Y_{ji}}{\partial \Lambda_{ji}} \cdot \frac{\partial \Lambda_{ji}}{\partial W_{(j-1)j}(r,s)} \quad (۹)$$

در رابطهٔ (۹)، Λ_j مطابق رابطهٔ (۱۰) خروجی شبکه در لایهٔ jام قبل از اعمال تابع فعالیت و Y_j بعد از اعمال آن است. B نیز بردار سطح آستانه را نشان می‌دهد.

$$\Lambda_j = Y_{j-1} W_{(j-1)j} - B \quad (۱۰)$$

همان‌طور که در رابطهٔ (۱۱) نشان داده شده است، تابع فعالیت برای لایهٔ خروجی خطی و برای سایر لایه‌ها غیرخطی سیگموید تعریف می‌شود.

$$Y_{ji} = \begin{cases} \dfrac{1}{1+\exp(-\Lambda_{ji})} & j = 2,3,\dots,o-1 \\ \Lambda_{ji} & j = o \end{cases} \quad (۱۱)$$

$$\frac{\partial Y_{ji}}{\partial \Lambda_{ji}} = \begin{cases} \Lambda_{ji}(1-\Lambda_{ji}) & j < o \\ 1 & j = o \end{cases} \quad (۱۲)$$

$$\delta_{ji}^{O} = -\frac{\partial E^{O}}{\partial \Lambda_{ji}} = -\frac{\partial E^{O}}{\partial Y_{ji}} \cdot \frac{\partial Y_{ji}}{\partial \Lambda_{ji}} = \begin{cases} (D_i - Y_{oi}) & j = o \\ Y_{ji}(1-Y_{ji}) \sum_{r=1}^{n_{j+1}} \delta_{(j+1)r}^{O} W'_{j(j+1)}(r,i) & j < o \end{cases} \quad (۱۳)$$

$$\delta_{ji}^{A} = -\frac{\partial E^{A}}{\partial \Lambda_{ji}} = -\frac{\partial E^{A}}{\partial Y_{ji}} \cdot \frac{\partial Y_{ji}}{\partial \Lambda_{ji}} = \begin{cases} Y_{ci}(1-Y_{ci})\left(M_{lai}^{A} - Y_{ci}\right) & j = c \\ Y_{ji}(1-Y_{ji}) \sum_{r=1}^{n_{j+1}} \delta_{(j+1)r}^{A} W'_{j(j+1)}(r,i) & j < c \end{cases} \quad (۱۴)$$

$$\delta_{ji}^{B}=-\frac{\partial E^{B}}{\partial \Lambda_{ji}}=-\frac{\partial E^{B}}{\partial Y_{ji}}\cdot\frac{\partial Y_{ji}}{\partial \Lambda_{ji}}=\begin{cases}Y_{c(i+n_A)}\left(1-Y_{c(i+n_A)}\right)\left(M_{lbi}^{B}-Y_{c(i+n_A)}\right) & j=c\\ Y_{ji}\left(1-Y_{ji}\right)\sum_{i=1}^{n_{j+1}}\delta_{(j+1)r}^{B}W'_{j(j+1)}(r,i) & j<c\end{cases} \quad (۱۵)$$

$$\begin{cases}\Delta W_{(j-1)j}=-\eta Y'_{j-1}\delta_{j}^{O} & j>c\\ \Delta W_{(c-1)c}(:,1:n_A)=-\eta Y'_{c-1}\left(\delta_{c(1:n_A)}^{O}+\delta_{c}^{A}\right) & \\ \Delta W_{(c-1)c}(:,n_A+1:n_A+n_B)=-\eta Y'_{c-1}\left(\delta_{c(n_A+1:n_A+n_B)}^{O}+\delta_{c}^{B}\right) & \\ \Delta W_{(j-1)j}=-\eta Y'_{j-1}\left(\delta_{j}^{O}+\delta_{j}^{A}+\delta_{j}^{B}\right) & j<c\end{cases} \quad (۱۶)$$

به‌منظور تعیین مقادیر لازم جهت اصلاح وزن‌های شبکه، پارامترهای δ_j^O (خطای پس‌انتشارشده از خروجی)، δ_j^A (خطای پس‌انتشارشده از بخش استخراج منیفلد A) و δ_j^B (خطای پس‌انتشار شده از بخش استخراج منیفلد B) تعریف می‌شوند. این پارامترها به‌عبارتی سیگنال‌های خطایی هستند که در شبکه به‌منظور اصلاح وزن‌ها پس‌انتشار می‌شوند. c اندیس لایهٔ تفکیک‌کننده منیفلدهاست. وزن‌های شبکه توسط رابطهٔ (۱۶) بروز می‌شوند که در این رابطه $Y_1 = X$ یعنی برابر با ورودی است.

۳- بهبود تفکیک درون‌منیفلدی

همان‌طور که عنوان شد خطاهای شکل‌گیری منیفلدها که در روابط (۲) و (۳) تعریف شده‌اند در جهت هم‌گراکردن نگاشت نمونه‌های هم‌طبقه در فضای منیفلد مربوطه عمل می‌کنند و در آنها معیاری برای ارزیابی نگاشت نمونه‌های غیرهم‌طبقه و افزایش پراکندگی برون‌طبقه‌ای لحاظ نشده است؛ لذا غالب‌شدن خطای شکل‌گیری منیفلدها به خطای پس‌انتشارشده از لایهٔ خروجی، منجر به استخراج مؤلفه‌های یکسان برای نمونه‌های غیرهم‌طبقه می‌شود. به‌خصوص اگر شباهت‌های برون‌طبقه‌ای به درون‌طبقه‌ای بیشتر باشد، این مسأله به‌طور کامل مشهود می‌شود. در این بخش برای حل این مسأله برای شکل‌گیری هر یک از منیفلدها بخش دیگری به تابع هزینه افزوده می‌شود که در جهت تمایز بیشتر مؤلفه‌های یادگیری‌شده برای هر طبقه در فضای هر منیفلد عمل می‌کند. روابط (۱۷) و (۱۸) این توابع را برای هریک از منیفلدها نشان می‌دهند. این توابع هزینه، توابع نمایی می‌باشند که با افزایش فاصله، مؤلفه‌های یادگیری‌شده برای هر طبقه در درون هر منیفلد، کمینه می‌شوند.

$$E_1^{MA}=\frac{1}{2}e^{-\sum_{i=1}^{n_A}\sum_{\substack{k=1\\k\neq la}}^{m_A}\left(M_{lai}^{A}-M_{ki}^{A}\right)^2} \quad (۱۷)$$

$$E_1^{MB}=\frac{1}{2}e^{-\sum_{i=1}^{n_B}\sum_{\substack{k=1\\k\neq lb}}^{m_B}\left(M_{lbi}^{B}-M_{ki}^{B}\right)^2} \quad (۱۸)$$

در این روابط m_A و m_B تعداد نقاط تعریف‌شده روی هر منیفلد هستند که معادل با تعداد طبقات در هر زیرفضا می‌باشند. δ_j^{MA} (خطای پس‌انتشارشدهٔ ناشی از تابع معیار E^{MA}) و δ_j^{MB} (خطای پس‌انتشارشدهٔ ناشی از تابع معیار E^{MB}) در روابط (۱۹) و (۲۰) تعریف شده‌اند.

با دقت در روابط (۱۹) و (۲۰) به‌نظر می‌رسد بتوان توابع هزینهٔ تعریف‌شده را به‌نحوی اصلاح کرد که مؤلفه‌های متمایزتری استخراج شوند. توابع بهبودیافته در (۲۱) و (۲۲) آورده شده‌اند.

$$\delta_{ji}^{MA}=-\frac{\partial E^{MA}}{\partial \Lambda_{ji}}=\begin{cases}2(1-\gamma)Y_{ci}(1-Y_{ci})E_1^{MA}\sum_{\substack{k=1\\k\neq la}}^{m_A}\left(M_{lai}^{A}-M_{ki}^{A}\right) & j=c\\ Y_{ji}\left(1-Y_{ji}\right)\sum_{r=1}^{n_{j+1}}\delta_{(j+1)r}^{MA}W'_{j(j+1)}(r,i) & j<c\end{cases} \quad (۱۹)$$

$$\delta_{ji}^{MB}=-\frac{\partial E^{MB}}{\partial \Lambda_{ji}}=\begin{cases}2(1-\gamma)Y_{c(i+n_A)}\left(1-Y_{c(i+n_A)}\right)E_1^{MB}\sum_{\substack{k=1\\k\neq lb}}^{m_B}\left(M_{lbi}^{B}-M_{ki}^{B}\right) & j=c\\ Y_{ji}\left(1-Y_{ji}\right)\sum_{r=1}^{n_{j+1}}\delta_{(j+1)r}^{MB}W'_{j(j+1)}(r,i) & j<c\end{cases} \quad (۲۰)$$

هستند. $E_1^{MA}=\frac{1}{2}e^{-\sum_{i=1}^{n_A}\sum_{\substack{k=1\\k\neq la}}^{m_A}\left(M_{lai}^{A}-M_{ki}^{A}\right)^2}$ از خروجی تمام نورون‌ها در این بخش ناشی می‌شود درحالی‌که $\alpha e^{-\sum_{\substack{k=1\\k\neq la}}^{m_A}\left(M_{lai}^{A}-M_{ki}^{A}\right)^2}$ مربوط به خروجی همان یک نورون است که δ_{ci}^{MA} برای آن محاسبه می‌شود. بنابراین مؤلفه‌های استخراج‌شده در خروجی هر نورون متمایزتر خواهند شد؛ اما در (۱۹) ممکن است در خروجی برخی نورون‌ها این تمایزها زیاد و برای برخی دیگر اندک شوند. در ارزیابی‌ها NMSNN1 و NMSNN2 مدل‌هایی هستند که به‌ترتیب توابع هزینه (۱۷) و (۱۸) و حالت بهبودیافتهٔ آنها به توابع هزینه مدل NMSNN0 افزوده شده است و برای آنها رابطهٔ (۱۶) به‌صورت (۲۵) اصلاح شده است.

$$E_2^{MA}=\frac{1}{2}\sum_{i=1}^{n_A}e^{-\sum_{\substack{k=1\\k\neq la}}^{m_A}\left(M_{lai}^{A}-M_{ki}^{A}\right)^2} \quad (۲۱)$$

$$E_2^{MB}=\frac{1}{2}\sum_{i=1}^{n_B}e^{-\sum_{\substack{k=1\\k\neq lb}}^{m_B}\left(M_{lbi}^{B}-M_{ki}^{B}\right)^2} \quad (۲۲)$$

برای این توابع δ_{ci}^{MA} و δ_{ci}^{MB} در روابط (۲۳) و (۲۴) اصلاح می‌شوند که در آنها α یک ضریب ثابت است. در مقایسه روابط (۱۹) و (۲۳) مشاهده می‌شود که این دو خطا در دو مقدار E_1^{MA} و $\alpha e^{-\sum_{\substack{k=1\\k\neq la}}^{m_A}\left(M_{lai}^{A}-M_{ki}^{A}\right)^2}$ باهم متفاوت

$$\delta_{ci}^{MA}=\alpha(1-\gamma)Y_{ci}(1-Y_{ci})e^{-\sum_{\substack{k=1\\k\neq la}}^{m_A}\left(M_{lai}^{A}-M_{ki}^{A}\right)^2}\sum_{\substack{k=1\\k\neq la}}^{m_A}\left(M_{lai}^{A}-M_{ki}^{A}\right) \quad (۲۳)$$

$$\delta_{ci}^{MB}=\alpha(1-\gamma)Y_{c(i+n_A)}\left(1-Y_{c(i+n_A)}\right)e^{-\sum_{\substack{k=1\\k\neq lb}}^{m_B}\left(M_{lbi}^{B}-M_{ki}^{B}\right)^2}\sum_{\substack{k=1\\k\neq lb}}^{m_B}\left(M_{lbi}^{B}-M_{ki}^{B}\right) \quad (۲۴)$$

$$\begin{cases}\Delta W_{(j-1)j}=-\eta Y'_{j-1}\delta_j^{O} & j>c\\ \Delta W_{(c-1)c}(:,1:n_A)=-\eta Y'_{c-1}\left(\delta_{c(1:n_A)}^{O}+\delta_c^{A}+\delta_c^{MA}\right)\\ \Delta W_{(c-1)c}(:,n_A+1:n_A+n_B)=-\eta Y'_{c-1}\left(\delta_{c(n_A+1:n_A+n_B)}^{O}+\delta_c^{B}+\delta_c^{MB}\right)\\ \Delta W_{(j-1)j}=-\eta Y'_{j-1}\left(\delta_j^{O}+\delta_j^{A}+\delta_j^{B}+\delta_j^{MA}+\delta_j^{MB}\right) & j<c\end{cases} \quad (۲۵)$$

۴- پیش‌تعلیم NMSNN

تعلیم شبکه‌های عصبی عمیق به‌دلیل مواجهه با تعداد بالای کمینه‌های موضعی، اغلب هم‌گرا نمی‌شود (Plath, Müller et al. 2008, Bengio 2009, Erhan, Bengio et al. 2010, Bengio 2012). درحالی‌که با مقداردهی اولیه مناسب وزن‌های شبکه به‌جای مقادیر تصادفی در ابتدای مسیر تعلیم، از بسیاری از این کمینه‌های موضعی می‌توان اجتناب کرد (Erhan, Manzagol et al. 2009, Erhan, Bengio et al. 2010, Glorot and Bengio 2010). مدل NMSNN نیز به‌عنوان یک شبکهٔ عصبی عمیق از این مقوله مستثنی نیست و شروع تعلیم آن با مقداردهی اولیه تصادفی، منجر می‌شود که تعلیم به یک کمینه موضعی ختم شود. درنتیجه نمی‌توان از تمام قابلیت شبکه برای تفکیک منیفلدها استفاده کرد. به همین دلیل برای تعلیم آن نیاز به بهره‌گیری از روش‌های پیش‌تعلیم است. روشی که در این مقاله مورد استفاده قرار گرفته روش پیش‌تعلیم لایه‌به‌لایه (سیدصالحی و سیدصالحی، ۱۳۹۱) است که در مقایسه با دو روش پیش‌تعلیم موجود، روش مشرف به هدف (نژادقلی، ۱۳۸۳؛ نژادقلی، ۱۳۹۱) و تجزیه به ماشین‌های بولتزمان (Hinton

and Salakhutdinov, 2006)، کارآیی بالاتری ارائه داده است (Seyyedsalehi and Seyyedsalehi, 2015).

در این روش شبکهٔ عصبی چندلایهٔ عمیق به تعداد متناظری شبکه با یک لایهٔ پنهان شکسته می‌شود و ابتدا این شبکه‌های یک‌لایهٔ پنهان، تعلیم داده می‌شوند؛ سپس مقادیر وزن حاصل از تعلیم اینها در شبکهٔ عصبی اصلی قرار داده می‌شود و برای تنظیم دقیق وزن‌ها، تعلیم یک‌پارچه صورت می‌گیرد (سیدصالحی و سیدصالحی، ۱۳۹۲).

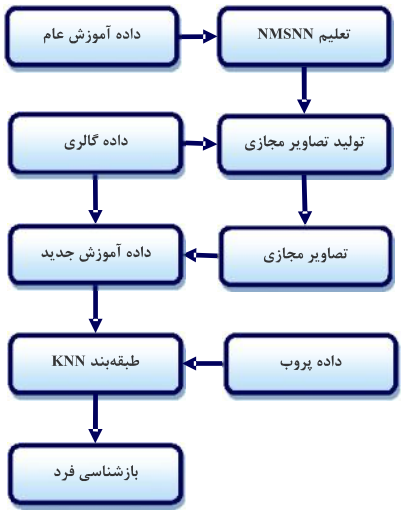
استفاده از پیش‌تعلیم لایه‌به‌لایه نه تنها منجر به هم‌گرایی تعلیم به کمینه مطلوب می‌شود، بلکه به‌طور چشم‌گیری سرعت تعلیم را بهبود می‌دهد؛ لذا مبتنی بر این روش، ابتدا مقادیر اولیهٔ وزن‌های دو بخش استخراج‌کنندهٔ مؤلفه‌های سطح پایین و بازسازی‌کننده تعیین می‌شود؛ سپس تعلیم یک‌پارچهٔ وزن‌های این دو بخش مانند یک شبکهٔ BNN عمیق صورت می‌گیرد. در ادامه با استفاده از مؤلفه‌های سطح پایین استخراج‌شده، مقادیر اولیهٔ وزن‌های بخش استخراج‌کنندهٔ مؤلفه‌های ذاتی تعیین و درنهایت تعلیم یک‌پارچهٔ کل ساختار انجام می‌شود.

۵- سامانهٔ بازشناس فرد

مسأله‌ای که در این مقاله مد نظر قرار گرفته است، بهبود بازشناسی چهره با یک تصویر از هر فرد است. در این نوع بازشناسی، از هر شخص فقط یک تصویر خنثی برای تعلیم مدل بازشناس فرد موجود است. درحالی‌که در هنگام استفاده از مدل، ممکن است لازم باشد تنوعات مختلفی از چهرهٔ این فرد مورد بازشناسی قرار گیرند. با اینکه امروزه ابزارهای بسیار دقیق و نیرومندی برای بازشناسی چهره موجود است، به نظر می‌رسد که کارآیی بسیاری از آنها برای این نوع مسأله به میزان زیادی افت می‌کند. غنی‌سازی دادگان تعلیم مدل بازشناس یکی از راه‌حل‌های مطرح برای برخورد با این مسأله است (Tan, Chen et al., 2006; Abdolali and Seyyedsalehi 2012; Mohammadzade and Hatzinakos, 2013).

در این مقاله با توجه با اینکه مدل‌های تعریف‌شده امکان تخمین مجازی حالت‌های دیگر فرد را فراهم کرده‌اند، از راه‌کار تولید نمونه‌های مجازی برای غنی‌سازی دادگان تعلیم طبقه‌بند استفاده شده است. بدین معنا که می‌توان برای چهرهٔ افراد جدیدی که در تعلیم مدل حضور نداشته‌اند، تنوعات مختلف را تولید و از این تصاویر مجازی برای غنی‌سازی دادگان تعلیم طبقه‌بند استفاده کرد.

در اینجا برای مقایسه و ارزیابی کارآیی روش پیشنهادی در این کاربرد، از یک طبقه‌بند مبتنی بر نزدیک‌ترین همسایگی (KNN[1]) با (K = 1) استفاده شده است. به این ترتیب که کارآیی آن در طبقه‌بندی تنوعات چهره، درحالتی‌که تنها یک چهره از هر فرد در دسترس است و حالت پایگاه داده غنی شده، با هم مقایسه شده است. در حالت اول داده آموزش KNN تصاویر گالری هستند که شامل یک تصویر خنثی از هر فرد می‌باشند، در مرحله بعد دادهٔ آموزش KNN توسط تصاویر مجازی غنی می‌شود.



(شکل ۵-۱): نمودار جریانی سامانهٔ بازشناس فرد با یک تصویر از هر فرد

نمودار جریانی این روش در (شکل ۵-۱) نشان داده شده است. همان‌طور که مشاهده می‌شود برای هر یک از دادگان، مدل NMSNN با داده آموزش عام تعلیم داده می‌شود و منیفلدهای مربوطه استخراج و با استفاده از اطلاعات هویت تصاویر گالری و مدل NMSNN تصاویر مجازی برای این افراد در حالت‌های مجازی بازسازی می‌شود؛ سپس این تصاویر مجازی در کنار تصاویر گالری به‌عنوان دادهٔ آموزش جدید KNN به‌کار می‌روند. مقدار حاصل از این ارزیابی به‌عنوان نرخ بازشناسی هویت[2] (IRR%) گزارش می‌شود.

[1] K-nearest Neighbors
[2] Identity Recognition Rate

۵-۱- تولید حالت‌های مجازی

منیفلد حالت‌های استخراج‌شده توسط مدل حاوی مؤلفه‌های مرتبط با حالت احساسی افراد است؛ لذا با حرکت روی این منیفلد می‌توان با حفظ هویت فرد، حالت یا وضعیت آن را تغییر داد. برای این منظور در مرحله بازترکیب منیفلدها اطلاعات حالت یا وضعیت آن مطابق رابطه (۲۶) با مکان مورد نظر روی منیفلد حالت جایگزین و تصاویری با حالت مجازی تولید می‌شود.

$$Y_c=\left[Y_{c(1:n_A)} \quad M_{lb}^{B}\right] \qquad (۲۶)$$

۵-۲- ارزیابی کیفیت حالت‌های مجازی

در این ارزیابی سعی شده است کیفیت تصاویر با حالت‌های مجازی تولید شده بررسی شود؛ بدین‌منظور از یک شبکهٔ عصبی با دو لایهٔ پنهان (۵۰ و ۲۰ نورون در لایه‌های پنهان) به‌عنوان شبکهٔ عصبی بازشناس حالت استفاده شده است. در خروجی این شبکه به تعداد حالت‌ها نورون قرار داده می‌شود که یک‌بودن هر یک نشان‌دهندهٔ یک حالت است. این شبکه ابتدا با تصاویر آموزش عام و پروب تعلیم داده می‌شود؛ سپس تصاویر با حالت‌های مجازی به شبکه داده می‌شود و برچسب حالت آن‌ها تخمین زده می‌شود. شایان ذکر است که علت اینکه برای این بخش مانند منیفلد افراد از KNN استفاده نشده، این است که KNN برچسب نمونه آزمون را براساس برچسب نزدیک‌ترین همسایه تعیین می‌کند و از آنجاکه شباهت‌های ناشی از یکی‌بودن فرد نسبت به یکی‌بودن حالت بیش‌تر است، اغلب نزدیک‌ترین همسایه براساس یکی‌بودن برچسب فرد تعیین می‌شود؛ لذا برای این بخش از شبکهٔ عصبی به‌عنوان طبقه‌بند استفاده شده است. مقدار حاصل از این بازشناسی برای بازشناسی حالت به‌عنوان نرخ بازشناسی حالت احساسی[1] (ERR%) گزارش می‌شود.

۶- پیاده‌سازی و نتایج

۶-۱- پایگاه داده

در این مقاله از دادگان CK+ (Lucey, Cohn et al. 2010) استفاده شده که توسعه‌یافته دادگان کوهن-کند[2] است. در پایگاه داده کوهن-کند دنباله‌ای از تصاویر برای شش حالت احساسی فرد موجود است که از حالت خنثی تا حالت موردنظر تغییر می‌کنند.

این حالت‌های احساسی شامل حالت‌های شادی، تعجب، خشم، ترس، تنفر و غم می‌باشند. همچنین برای هر تصویر چهره، برداری از نشانه‌ها وجود دارد که می‌توان از آن برای تراز‌کردن چهره‌ها استفاده کرد. در این مقاله چهره‌ها طوری تراز شده‌اند که چشم‌ها در یک راستای افقی قرار بگیرند (محمدیان و همکاران، ۱۳۹۲). اطلاعات اضافه اطراف تصاویر نیز به‌نحوی حذف شده است که همهٔ جزییات مهم در تصویر موجود باشند یا به‌عبارت دیگر تصویر فقط حاوی چهره باشد. درنهایت همهٔ تصاویر به اندازهٔ ۵۰×۵۰ تبدیل شدند. در (شکل ۶-۱) نمونه‌هایی از تصاویر آماده‌شده از این پایگاه داده برای حالت‌های احساسی مورد استفاده آورده شده‌اند؛ هم‌چنین تصاویر نیمه اول هر یک از این دسته‌ها حذف شدند، تا تصاویری که به‌طور واقعی نمایان‌گر حالت احساسی مربوطه هستند، نگه داشته شوند.

غم شادی ترس تنفر تعجب خشم خنثی

(شکل ۶-۱): نمونه‌هایی از تصاویر پایگاه داده کوهن-کند برای حالت‌های احساسی مختلف بعد از ترازنمودن چهره‌ها و حذف اطلاعات اضافه اطراف تصاویر

تصاویر مربوط به ده هویت مختلف به‌طور تصادفی برای آزمون و تصاویر مربوط به سایر هویت‌ها نیز تحت عنوان تصاویر آموزش عام برای استخراج منیفلدهای مربوطه به‌کار رفته‌اند. تصاویر آزمون برای ارزیابی مدل‌های استخراج‌کننده منیفلد جدا شده و در تعلیم آنها نقشی نداشته‌اند. همان‌طور که در (شکل ۶-۲) نشان داده شده است، تصاویر آزمون نیز به دو بخش گالری و پروب تقسیم شده‌اند. مجموعه گالری شامل یک تصویر خنثی از هر فرد آزمون است که برای تولید تصاویر مجازی به‌کار رفته است. مجموعه پروب نیز مربوط به سایر تصاویر افراد آزمون است که برای ارزیابی کیفیت تصاویر آزمون استفاده شده است.

(شکل ۶-۲): روش تقسیم‌بندی دادگان

[1] Expression Recognition Rate
[2] Cohn-Kanade

برای این دادگان با توجه به ماهیت آن دو منیفلد هویت و حالت احساسی به‌عنوان منیفلدهای A و B به‌ترتیب شامل ۹۶ هویت و هفت حالت احساسی مختلف فرض شده است.

۶-۲- مدل NMSNN0

ساختار NMSNN0 برای یادگیری منیفلدهای مفروض مطابق (جدول ۶-۱) در نظر گرفته شده است. مطابق این جدول برای یادگیری منیفلدهای حالت و هویت به‌ترتیب ۷ و ۹۶ نورون فرض شده است که البته می‌توان مقادیر دیگری نیز اختیار کرد؛ اما باید توجه داشت که تعداد آنها باید به‌نحوی انتخاب شود که قادر به یادگیری ۷ حالت مختلف و همچنین ۹۶ هویت متفاوت باشند.

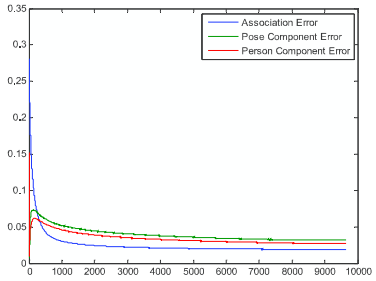
برای ارزیابی تأثیر استفاده از روش پیش‌تعلیم لایه‌به‌لایه برای مقداردهی اولیه NMSNN0، در گام اول این شبکه مطابق (داداشی و همکاران، ۱۳۹۰) به‌صورت تصادفی مقداردهی شد و تعلیم آن صورت گرفت. در گام بعد با استفاده از این روش پیش‌تعلیم، به مقداردهی اولیه وزن‌ها پرداخته شد.

(جدول ۶-۱): پارامترهای مدل NMSNN0

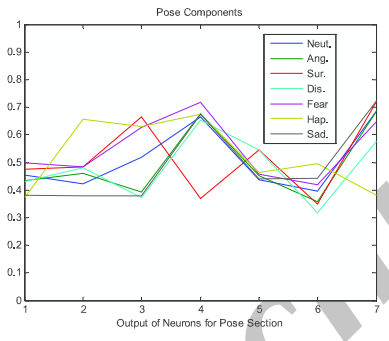
پارامتر	مقدار یا نوع
تعداد لایه‌های پنهان	۷
تعداد نورون‌های لایه‌های پنهان	۱۰۰۰-۴۰۰-۱۰۰-۹۶+۷-۱۰۰-۴۰۰-۱۰۰۰
تابع نورون‌های لایه پنهان	غیرخطی
تابع نورون‌های خروجی	خطی
ضریب یادگیری	۰/۰۰۰۱ -۰/۰۰۱
ضریب گشتاور	۰/۷
شرط توقف یادگیری	توسط کاربر

در (شکل ۶-۳) نمودار خطای تعلیم خودانجمنی و خطای شکل‌گیری مؤلفه‌های هویت و حالت در یادگیری منیفلدهای مربوطه در مرحله پیش‌تعلیم بخش استخراج‌کننده مؤلفه‌های ذاتی رسم شده است.

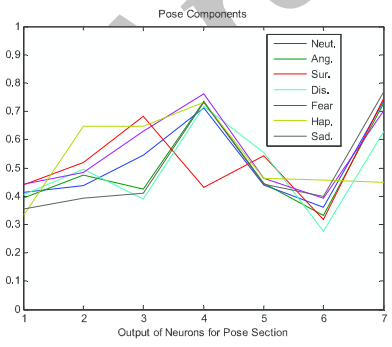
همان‌طور که مشاهده می‌شود، منحنی هر سه خطا به‌خوبی همگرا شده و به کمینه مطلوبی رسیده است. معیار برای توقف یادگیری ثابت‌شدن مقادیر این خطاها بوده است. مقادیر وزن حاصل از این مرحله در NMSNN0 قرار داده شد و تعلیم یک‌پارچه مدل صورت گرفت. نمودارهای (شکل ۶-۴) مؤلفه‌های اساسی استخراج‌شده را برای هفت حالت مختلف نشان می‌دهد. همان‌طور که مشاهده می‌شود، قبل و بعد از تعلیم یک‌پارچه تفاوت معناداری بین این دو نمودار وجود ندارد. به‌عبارت دیگر در طی مراحل پیش‌تعلیم، آموزش اصلی مدل صورت گرفته است. با توجه به این نمودارها می‌توان عنوان کرد برای هفت حالت، هفت مؤلفهٔ متمایز شکل گرفته که این تمایز برای دو حالت تعجب و شادی مشهودتر است. چون برای این دو حالت، تغییر چهره نسبت به سایر حالت‌ها شدیدتر است.



(شکل ۶-۳): نمودار خطای تعلیم خودانجمنی و خطای شکل‌گیری مؤلفه‌های هویت و حالت در طی پیش‌تعلیم بخش استخراج‌کننده مؤلفه‌های ذاتی NMSNN0



الف



ب

(شکل ۶-۴): نمودار مؤلفه‌های اساسی استخراج‌شده برای هفت حالت احساسی توسط NMSNN0، نمودارهای الف و ب به‌ترتیب قبل و بعد از تعلیم یک‌پارچه

با استفاده از منیفلد حالت استخراج‌شده برای تصاویر گالری مطابق روشی که در بخش (۵-۱) عنوان شد، حالت‌های احساسی مجازی بازسازی شد که (شکل ۶-۵) این تصاویر را برای یکی از افراد آزمون نشان می‌دهد. ستون اول تصویر خنثی واقعی است که به مدل NMSNN0 داده شده است و ستون‌های بعدی تصاویر مجازی مربوط به حالت‌های احساسی بازسازی‌شده را نشان می‌دهند. سطر اول تصاویر مجازی بازسازی‌شده توسط NMSNN0 بدون پیش‌تعلیم، NMSNN0 پیش‌تعلیم‌یافته بدون تعلیم یک‌پارچه و NMSNN0 را بعد از تعلیم یک‌پارچه نشان می‌دهند. در مقایسهٔ تصاویر سطر اول با سایر تصاویر مشاهده می‌شود که حالت تصاویر مجازی بازسازی‌شده تاحدودی خنثی است که این نشان‌دهنده این است که مدل NMSNN0 بدون پیش‌تعلیم قادر به استخراج منیفلد حالت از دادگان CK+ نبوده است. همچنین اطلاعات هویت تصویر تخریب شده است، به‌گونه‌ای‌که هویت تصایر بازسازی‌شده تا حدودی از تصویر واقعی فاصله گرفته است.

(شکل ۶-۵): ستون اول تصویر خنثی واقعی یکی از افراد آزمون را نشان می‌دهد که به مدل NMSNN0 داده شده است و ستون‌های بعدی تصاویر مجازی مربوط به حالت‌های احساسی بازسازی‌شده را نشان می‌دهند. سطرهای اول تا سوم به‌ترتیب توسط NMSNN0 بدون پیش‌تعلیم، NMSNN0 پیش‌تعلیم‌یافته بدون تعلیم یک‌پارچه و NMSNN0 بعد از تعلیم یک‌پارچه.

برای مقایسهٔ کمّی تأثیر پیش‌تعلیم لایه‌به‌لایه در یادگیری بهتر منیفلدها توسط NMSNN0، کیفیت تصاویر مجازی تولیدشده با روش‌های عنوان‌شده در بخش (۵) مورد ارزیابی قرار گرفت که نتایج آن در (جدول ۶-۲) آورده شده است. ستون دوم در این جدول نرخ بازشناسی حالت را برای تصاویر مجازی تولیدشده توسط مدل‌ها را با استفاده از شبکهٔ عصبی بازشناس حالت نشان می‌دهد. ستون سوم نیز تأثیر غنی‌سازی دادگان تعلیم KNN را در بهبود نرخ بازشناسی هویت نشان می‌دهد.

این نتایج تأیید می‌کند که روش پیش‌تعلیم لایه‌به‌لایه نقش به‌سزایی در یادگیری مدل دارد. این روش با تعلیم لایه‌به‌لایهٔ مدل در مراحل استخراج مؤلفه‌های سطح پایین، در استخراج مؤلفه‌های معنادارتر کمک می‌کند. هرچه این مؤلفه‌ها حاوی اطلاعات بیشتری باشند، یادگیری منیفلدها در بخش استخراج‌کنندهٔ مؤلفه‌های ذاتی بهتر صورت می‌گیرد که نتایج این جدول نیز این را نشان می‌دهند.

(جدول ۶-۲): مقایسه تأثیر پیش‌تعلیم در بهبود بازشناسی هویت تصاویر پروب و حالت تصاویر مجازی تولیدشده توسط مدل‌ها برای دادگان CK+.

مدل تولیدکننده تصاویر مجازی	ERR%	IRR%
---	---	۸۸/۵۶٪
NMSNN0 بدون پیش‌تعلیم	۲۴/۲۹٪	۹۰/۶۲٪
NMSNN0 قبل از تعلیم یکپارچه	۷۰٪	۹۷/۰۷٪
NMSNN0 بعد از تعلیم یکپارچه	۷۱/۴۳٪	۹۷/۰۷٪

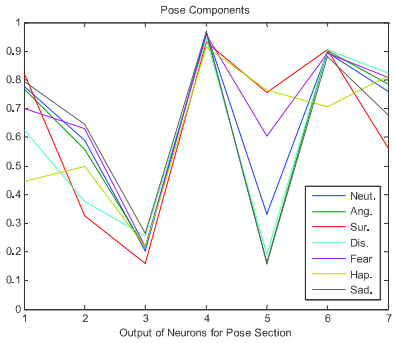
۶-۳- اصلاح تابع هزینه با هدف افزایش تمایز بین طبقات

در این بخش به ارزیابی مدل‌های NMSNN1 و NMSNN2 که در آنها سعی شده است مؤلفه‌های استخراج‌شده برای طبقات مختلف با بهره‌گیری از توابع هزینه تعریف‌شده متمایزتر شوند، پرداخته شده است. دادگان و نحوهٔ آزمایش مانند بخش قبل است با این تفاوت که اصلاح وزن‌ها مطابق با رابطه (۲۵) صورت می‌گیرد. در نمودارهای (شکل ۶-۶)

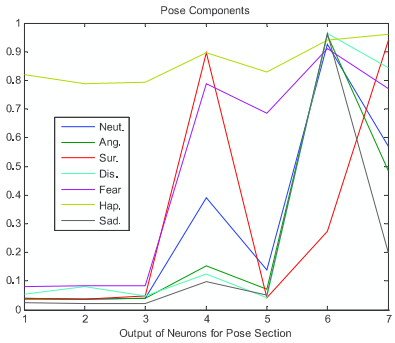
مؤلفه‌های اساسی استخراج‌شده برای هفت حالت احساسی در دادگان CK+ به‌ترتیب توسط مدل‌های NMSNN1 و NMSNN2 آورده شده است. با توجه به اینکه تعداد مؤلفه‌های حالت کمتر و تأثیر توابع هزینهٔ افزوده‌شده بهتر قابل مشاهده است، فقط مؤلفه‌های حالت رسم شده‌اند. در مقایسه نمودارهای الف در (شکل ۶-۴) و (شکل ۶-۶) مشاهده می‌شود که دامنهٔ مؤلفه‌های یادگیری شده در NMSNN1 بیشتر و مقادیر این مؤلفه‌ها در خروجی برخی نورون‌ها متمایزتر می‌باشند. با این وجود برای نورون چهار در نمودار الف (شکل ۶-۶) مقادیر تمام مؤلفه‌ها یکسان است.

علت آن را باید در تابع هزینهٔ تعریف‌شده برای متمایزکردن مؤلفه‌ها جستجو کرد. همان‌طور که در بخش (۳) عنوان شد خطای پس‌انتشارشده ناشی از تابع هزینهٔ متمایزساز مؤلفه‌های به‌کار رفته در NMSNN1 به‌طور عمده برای هر نورون تحت تأثیر میزان تمایز مؤلفه‌ها در خروجی تمام نورون‌هاست؛ لذا تمایز زیاد در خروجی برخی نورون‌ها، شباهت آن‌ها را در خروجی دیگر نورون‌ها جبران می‌کند؛ این باعث می‌شود که مانند نمودار الف برای برخی نورون‌ها خروجی‌ها مشابه و برای برخی دیگر به میزان خوبی متمایز شوند. با اصلاح تابع هزینه در مدل NMSNN2 این مشکل حل شده است. نمودار ب (شکل ۶-۶) این مؤلفه‌ها را که توسط این مدل استخراج شده‌اند نشان می‌دهد. همان‌طور که مشاهده می‌شود، مؤلفه‌ها در خروجی تمام نورون‌ها متمایز هستند. در اینجا نیز مؤلفه‌های شکل‌گرفته برای حالت‌های شادی و تعجب که در آنها چهره از حالت خنثی فاصلهٔ بیشتری می‌گیرد، متمایزتر هستند.

برای ارزیابی کمّی نیز تصاویر مجازی با استفاده از این مدل‌ها مطابق آنچه که در بخش (۵) عنوان شد، تولید شد و مورد ارزیابی قرار گرفت که نتایج آن در (جدول ۶-۳) آورده شده است. نتایج این جدول نشان می‌دهد که با هدایت مدل در جهت استخراج مؤلفه‌های متمایزتر برای حالت‌ها و هویت‌های مختلف، مؤلفه‌های یادگیری‌شده به مؤلفه‌های ذاتی بیان‌گر هر حالت، نزدیک‌تر خواهند بود. به‌گونه‌ای‌که حالت تصاویر بازسازی‌شده با حالت‌های مجازی برای دادهٔ آزمون با استفاده از مؤلفه‌های ذاتی حالت، با حالت‌های موردنظر تطابق بیشتری دارند. افزایش ERR% از ۷۰٪ به ۷۵/۷۱٪ این مسأله را تأیید می‌کند.



الف



ب

(شکل ۶-۶): نمودار مؤلفه‌های اساسی استخراج‌شده برای هفت حالت احساسی در دادگان CK+، الف وب به‌ترتیب توسط NMSNN1 و NMSNN2

(جدول ۶-۳): ERR% در بازشناسی حالت تصاویر مجازی تولیدشده توسط مدل‌ها برای دادگان CK+.

مدل تولیدکننده تصاویر مجازی	ERR%
NMSNN0	۷۰٪
NMSNN1	۷۲/۸۶٪
NMSNN2	۷۵/۷۱٪

۷- بحث و نتیجه‌گیری

این مقاله به شرح کامل و توسعهٔ مدل ارائه‌شده در (داداشی و همکاران، ۱۳۹۰) پرداخته و نشان داده شده است که با استفاده از روش پیش‌تعلیم لایه‌به‌لایه (سیدصالحی و سیدصالحی، ۱۳۹۲)، قابلیت این مدل برای استخراج منیفلدهای تصاویر چهره به‌طور معناداری افزایش می‌یابد. چون با توجه به ساختار عمیق این مدل تعلیم آن با مقداردهی اولیه تصادفی به یک کمینه محلی ختم می‌شود

که در نتیجه بهره‌گیری از تمام قابلیت مدل را محدود می‌کند.

این ساختار برای بازشناسی چهره با یک تصویر از هر فرد با راه‌کار غنی‌سازی دادگان تعلیم از طریق تولید نمونه‌های مجازی مورد استفاده قرار گرفت؛ بدین طریق که با بهره‌گیری از منیفلد حالت استخراج‌شده برای هر تصویر خنثی آزمون، حالت‌های مختلف به‌صورت مجازی بازسازی شد. بهبود در نرخ بازشناسی هویت نیز، نتیجهٔ این غنی‌سازی بود.

هم‌چنین به‌منظور افزایش تمایز درون‌منیفلدی بین طبقات مختلف، بخش دیگری به تابع معیار یادگیری منیفلدها افزوده شده که در دو گام نیز بهبود داده شده است. به‌نظر می‌رسد وجود این بخش در مواردی‌که شباهت‌های برون‌طبقه‌ای به درون‌طبقه‌ای در فضای یک منیفلد بیشتر باشد، ضروریست.

مراجع

محمدیان، امین، آقایی‌نیا، حسن و توحیدخواه، فرزاد، "بازشناسی جلوه‌های هیجانی چهره مستقل از فرد مبتنی بر دانش اولیه از شخص جدید"، مجله مهندسی پزشکی زیستی، ۱۳۹۱، دوره ۶، شماره ۳، صفحات ۲۰۷-۲۱۸.

داداشی، ندا، عبدالعلی، فاطمه و سیّدصالحی، سیّد علی، "بهبود بازشناسی چهره با یک تصویر از هر فرد به روش تولید تصاویر مجازی توسط شبکه‌های عصبی"، دوفصل‌نامه علمی پژوهشی پردازش علائم و داده‌ها، ۱۳۹۰، شماره ۱، پیاپی ۱۵، صفحات ۳۳-۴۳.

داداشی, ندا، "بازشناسی چهره توسط یک چهره از هر فرد"، پایان‌نامه کارشناسی ارشد بیوالکتریک، دانشکدهٔ مهندسی پزشکی، دانشگاه صنعتی امیرکبیر، ۱۳۸۷.

سیّدصالحی، سیّده زهره و سیّدصالحی، سیّد علی، "روش پیش‌تعلیم سریع برای آموزش شبکه‌های عصبی با ساختار عمیق"، نوزدهمین کنفرانس مهندسی پزشکی ایران، ۱۳۹۱.

سیّدصالحی، سیّده زهره و سیّدصالحی، سیّد علی، "روش پیش‌تعلیم سریع بر مبنای کمینه‌سازی خطا برای همگرائی یادگیری شبکه‌های عصبی با ساختار عمیق"، دوفصل‌نامه علمی پژوهشی پردازش علائم و داده‌ها، ۱۳۹۲، شماره ۱، پیاپی ۱۹، صفحات ۱۳-۲۶.

عبدالعلی، فاطمه، "بهبود مدل تحلیل‌گر غیرخطی اطلاعات چهره به روش به‌کارگیری اتصالات بازگشتی و جاذب‌ها"، پایان‌نامه کارشناسی ارشد بیوالکتریک، دانشکدهٔ مهندسی پزشکی، دانشگاه صنعتی امیرکبیر، ۱۳۸۹.

نژادقلی، ایثار، "آزمایش‌هایی جهت دوسویه‌کردن شبکه‌های عصبی"، گزارش تحقیقاتی پژوهشکدهٔ پردازش هوشمند علائم، بخش پردازش گفتار، ۱۳۸۳.

نژادقلی، ایثار، "مدل‌سازی نحوهٔ استخراج و بازترکیب ویژگی‌های ادراکی در مغز با لحاظ‌کردن تعاملات بین آنها"، پایان‌نامهٔ دکترای بیوالکتریک، دانشکدهٔ مهندسی پزشکی، دانشگاه صنعتی امیرکبیر، ۱۳۹۱.

Abdolali, F. and Seyyedsalehi S.A., (2012), "Improving face recognition from a single image per person via virtual images produced by a bidirectional network", Procedia-Social and Behavioral Sciences 32: 108-116.

Bengio, Y., (2009), "Learning deep architectures for AI", Foundations and Trends® in Machine Learning 2(1): 1-127.

Bengio, Y., (2012), "Evolving culture vs local minima", arXiv preprint arXiv: 1203.2990.

Chen, S., Liu J. and Zhou Z.H., (2004), "Making FLDA applicable to face recognition with one sample per person", Pattern recognition 37(7): 1553-1555.

Erhan, D., Manzagol P.A., Bengio Y., Bengio S. and Vincent P., (2009), "The difficulty of training deep architectures and the effect of unsupervised pre-training", Proceedings of The Twelfth International Conference on Artificial Intelligence and Statistics (AISTATS'09), (Citeseer2009), pp. 153-160.

Erhan, D., Bengio Y., Courville A., Manzagol P.A., Vincent P. and Bengio S., (2010), "Why does unsupervised pre-training help deep learning?", The Journal of Machine Learning Research 11: 625-660.

Glorot, X. and Bengio Y., (2010), "Understanding the difficulty of training deep feedforward neural networks", Proceedings of the International Conference on Artificial Intelligence and Statistics (AIST-ATS'10). Society for Artificial Intelligence and Statistics.

Hinton, G.E. and Salakhutdinov R.R., (2006), "Reducing the dimensionality of data with neural networks", Science 313(5786): 504-507.

Hsieh, C.K., Lai S.H. and Chen Y.C., (2010), "An optical flow-based approach to robust face recognition under expression variations", Image Processing, IEEE Transactions on 19(1): 233-240.

Kan, M., Shan S., Su Y., Xu D. and Chen X., (2013), "Adaptive discriminant learning for face recognition", Pattern Recognition 46: 2497- 2509.

Lucey, P., Cohn J. F., Kanade T., Saragih J., Ambadar Z. and Matthews I., (2010), "The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression", 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).

Martínez, A. M., (2002), "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class", IEEE Transactions on Pattern Analysis and Machine Intelligence 24(6): 748-763.

Martınez, A. M. (2003). "Matching expression variant faces", Vision research 43(9): 1047-1060.

Mohammadzade, H. and Hatzinakos D., (2013), "Projection in to expression subspaces for face recognition from single sample per person", IEEE Transactions on Affective Computing 4(1): 69-82.

Plath, N., (2008), "Extracting low-dimensional features by means of deep network architectures", PhD. Thesis, Technische Universität Berlin.

Seow, M.J, (2006), "Learning as a nonlinear line of attraction for pattern association, classification and recognition", M.S. Thesis, Old Dominion University.

Seyyedsalehi S.Z., Seyyedsalehi S.A., (2014), "Simultaneous learning of nonlinear manifolds based on the bottleneck neural network", Neural Processing Letters 40, 191-209.

Seyyedsalehi S.Z. and Seyyedsalehi S.A. (2015), "A fast and efficient pre-training method based on Layer-by-layer maximum discrimination for deep neural networks", Neurocomputing, (In press).

Tan, X., Chen S., Zhou Z.H. and Zhang F., (2006). "Face recognition from a single image per person: A survey", Pattern Recognition 39(9): 1725-1745.

Wang, B., Li W., Li Z. and Liao Q., (2013), "Adaptive linear regression for single-sample face recognition", Neurocomputing 115: 186-191.

Zhu, N., Tang T., Tang S., Tang D. and Yu F, (2013), "A sparse representation method based on kernel and virtual samples for face recognition", International Journal for Light and Electron Optics 124(23): 6236-6241.

سیّده زهره سیّدصالحی مدرک کارشناسی خود را دررشتهٔ مهندسی پزشکی- بیوالکتریک از دانشگاه صنعتی امیرکبیر در سال ۱۳۸۳، کارشناسی ارشد را در همان رشته از دانشکدهٔ فنی دانشگاه شاهد در سال ۱۳۸۶ و دکترای خود را دررشتهٔ مهندسی پزشکی- بیوالکتریک از دانشگاه صنعتی امیرکبیر در سال ۱۳۹۲ دریافت کرده است. زمینه‌های پژوهشی مورد علاقهٔ ایشان یادگیری عمیق، شبکه‌های عصبی مصنوعی، پردازش خطی و غیرخطی سیگنال‌ها و روش‌های یادگیری منیفلد است.

نشانی رایانامهٔ ایشان عبارت است از:

z.seyyedsalehi@aut.ac.ir

سیّد علی سیّدصالحی مدرک کارشناسی خود را در رشتهٔ مهندسی برق از دانشگاه صنعتی شریف در سال ۱۳۶۱، کارشناسی ارشد را در رشتهٔ مهندسی برق از دانشگاه صنعتی امیرکبیر در سال ۱۳۶۷ و دکترای خود را در رشتهٔ مهندسی برق- بیوالکتریک از دانشگاه تربیت مدرس در سال ۱۳۷۴ دریافت کرده است. وی در حال حاضر دانشیار دانشکدهٔ مهندسی پزشکی دانشگاه صنعتی امیرکبیر است. زمینه‌های پژوهشی مورد علاقهٔ ایشان پردازش و بازشناسی گفتار، شبکه‌های عصبی مصنوعی و زیستی، مدل‌سازی عملکرد مغز و پردازش خطی و غیرخطی سیگنال است.

نشانی رایانامهٔ ایشان عبارت است از:

ssalehi@aut.ac.ir