# ارائه یک بستر یادگیری عمیق نیمه نظارتی برای بازسازی سه‌بعدی چهره از یک تصویر دوبعدی

**شیما کامیاب[۱]، سیده زهره عظیمی فر[۲]**
۱ - دانشجوی دکترا کامپیوتر، دانشگاه شیراز
shima.kamyab@gmail.com
۲ - دانشیار دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه شیراز
azimifar@cse.shirazu.ac.ir

**چکیده:** در این مقاله یک بستر یادگیری عمیق نیمه نظارتی برای بازسازی سه‌بعدی از یک تصویر دوبعدی پیشنهاد شده است که در آن به منظور کاهش نیاز به برچسب سه‌بعدی و دوبعدی از دو بخشِ بدون نظارتِ از پیش آموزش داده شده استفاده شده است. بدین ترتیب با بهره‌گیری از بخش‌های آموزش دیده، به منظور آموزش کل شبکه، به داده برچسب دار کمتری نیاز است، علاوه بر اینکه با توجه به استفاده از داده به عنوان تنها منبع دانش برای یادگیری، نیازی به استفاده از فرض‌های مختلف در مورد چگونگی شکل‌گیری تصویر نخواهد بود. ایده اصلی در بستر پیشنهادی، یافتن نگاشتی بین فضای‌های بازنمایی با ابعاد پایین‌تر دوبعدی و سه‌بعدی می‌باشد. بنابراین بستر پیشنهادی در این مقاله شامل بخش‌های بدون نظارتِ نگاشت از فضاهای دوبعدی و سه‌بعدی به بازنمایی‌های بعد پایین، و بخش نظارتی نگاشت بین بازنمایی‌های بعدپایین می‌باشد. نتایج ارزیابی و مقایسه بستر پیشنهادی با چند بستر مشابه موجود روی پایگاه‌های داده چهره‌ی انسان، نشان‌دهنده کارایی مطلوب بستر نیمه نظارتی پیشنهادی در بازسازی سه‌بعدی از یک تصویر دوبعدی است. این بستر می‌تواند قدمی مفید در جهت هوشمندسازی فعالیت نیروی انتظامی برای تشخیص چهره باشد.

**واژه‌های کلیدی:** بازسازی سه‌بعدی از یک تصویر دوبعدی، نگاشت بازنمایی دو بعدی به سه‌بعدی، بازسازی سه‌بعدی نیمه نظارتی. هوشمندسازی فعالیت ناجا با بازسازی سه‌بعدی، یادگیری عمیق در بازسازی سه‌بعدی.

**تاریخ ارسال مقاله:** ۹۹/۰۴/۰۵
**تاریخ پذیرش مقاله:** ۹۹/۰۷/۰۲
**نام نویسنده مسئول:** سیده زهره عظیمی فر

## ۱- مقدمه

بازسازی سه‌بعدی یکی از مسائل مطرح در حوزه‌ی یادگیری ماشین[1] و بینایی ماشین[2] می‌باشد که امروزه توجه زیادی را در کاربردهای مختلف به خود جلب کرده است [۱۹-۱]. علاوه براین بازسازی سه‌بعدی چهره انسان برای بهبود کیفیت و هوشمندسازی بازشناسی چهره در فعالیت های نیروی انتظامی می‌تواند بسیار مفید واقع شود.

ماهیت مسئله بازسازی سه‌بعدی از یک تصویر دوبعدی بدفرم[3] است با این ویژگی که برای یک تصویر دوبعدی تعداد نامتناهی شکل سه‌بعدی می‌توان یافت که به تصویر مورد نظر نگاشت شود [۷,۱۱]. این به این دلیل است که هدف در مسئله بازسازی سه‌بعدی، بازیابی بعد سوم از روی یک اندازه‌گیری دو بعدی است و در این موارد می‌توان تعداد نامحدود شکل سه‌بعدی غیر عملی یافت که به اندازه‌گیری موجود نگاشت شوند. در این حالت به اصطلاح یکتا بودن[4] جواب نقض می‌شود که یکی از شرط‌های هادامارد برای خوش فرم[5] بودن یک مسئله است [۲۰]. به منظور حل چنین مسائلی نیاز است که در مورد مسئله یکسری فرضیات اتخاذ شود تا فضای جستجو به نواحی قابل قبول محدود شود. به عنوان مثال می‌توان از یک توزیع احتمال روی اشکال سه‌بعدی به عنوان دانش اولیه استفاده کرد تا مشخص شود چه اشکالی با احتمال بیشتری جواب مورد قبول برای مسئله هستند [۱۱].

از دیدگاه یادگیری ماشین، روند حل مسئله بازسازی سه‌بعدی می‌تواند به سه بخش تقسیم شود:

- انتخاب مدل با توجه به تئوری "No free lunch" [۲۱] به منظور محدود کردن فضای جستجو. مانند استفاده از دانش بیشتر مانند چندین تصویر ورودی یا دنباله [۲۲,۱۹-۲۴]یا یک توزیع آماری [۲۵] یا استفاده از فرض‌های رگرسیون برای ارتباط دادن داده دو بعدی ورودی به شکل سه‌بعدی خروجی [۷, ۸, ۱۱-۱۶, ۲۶].
- تعریف یک معیار امتیازدهی به منظور هدایت روند جستجو در فضای تعریف شده با مدل مورد استفاده. به عنوان مثال در مسئله بازسازی سه‌بعدی معیار امتیازدهی طوری تعریف می‌شود که شکل یافته شده نه خیلی کلی[6] باشد که ویژگی های بارز را از دست دهد و نه خیلی بیش از نیاز حاوی جزییات باشد تا مدل بیش پردازش[7] شود [۱۶].
- تعریف یک استراتژی جستجو برای یافتن راه‌حل در فضای مورد نظر با معیار امتیازدهی تعریف شده. روش‌های جستجوی بسیاری تاکنون مطرح شده اند ازقبیل یادگیری‌های با نظارت، بدون نظارت و نیمه نظارتی. به عنوان مثال در بازسازی سه‌بعدی هم میتوان از یک مدل سه‌بعدی برای تولید برچسب برای یادگیری با نظارت بهره برد[۱] و هم می‌توان از یک ارائه‌دهنده[8] در انتهای بستر بازسازی سه‌بعدی بهره برد تا شکل سه‌بعدی تولید شده توسط شبکه را به تصویر دوبعدی تبدیل کند و یک یادگیری بدون نظارت را شکل داد [۲۵].

در این مقاله از از ساختارهای یادگیری عمیق به عنوان مدل مورد استفاده و روند آموزش نیمه نظارتی پس انتشار خطا به عنوان استراتژی جستجو و تابع هزینه شبکه عصبی که در اینجا مجموع مربعات خطا[9] (MSE) استفاده می‌شود، به عنوان معیار امتیازدهی در نظر گرفته شده است. در مورد مسئله بازسازی سه‌بعدی اگر بتوان فرض‌های مسئله را طوری تعریف کرد که روند شکل‌گیری تصویر بتواند به طور دقیق بیان شود آنگاه می‌توان برای حل آن یک جواب شکل بسته[10] یافت اما اتخاذ فرض‌های زیاد باعث می‌شود فضای جستجو بیش از حد محدود شود و بنابراین جواب‌های یافته شده مطلوب نباشند بنابراین از ساختارهای مبتنی بر یادگیری مانند شبکه های عمیق برای حل اینگونه مسائل بدفرم استفاده می‌شود تا از داده‌ی آموزش دانش مورد نیاز یادگرفته شود و نیازی به اتخاذ فرض‌های محدود کننده نباشد.

از سوی دیگر فراهم کردن داده موردنیاز برای آموزش شبکه‌های عمیق یکی از چالش‌های اساسی در حوزه یادگیری عمیق می‌باشد. به دلیل تعداد زیاد پارامترهای آزاد در شبکه‌های عمیق، برای آموزش مناسب این شبکه‌ها و جلوگیری از بیش پردازش شدن آن‌ها، نیاز به مجموعه داده‌های بزرگ برای آموزش می‌باشد. در حوزه بازسازی سه‌بعدی فراهم کردن داده زیاد واقعی با برچسب سه‌بعدی متناظر بسیار دشوار است و نیاز به ایزار

[1] Machine learning
[2] Computer vision
[3] Ill-posed
[4] Uniqueness
[5] Well-posed

[6] Generic
[7] Overfit
[8] Renderer
[9] Mean Squared Error (MSE)
[10] Closed form

خاص و زمان فراوان دارد. یک راه‌حل استفاده از ساختارهای تولیدگر[1] برای تولید داده مورد نیاز این شبکه‌ها است. ساختارهای تولیدگر ابزاری هستند که با استفاده از دانشی که از داده واقعی کسب می‌کنند می‌توانند داده مصنوعی با کیفیت قابل مقایسه با داده واقعی تولید کنند. اما داده مصنوعی قدرت شبکه را به داده‌های مصنوعی با روش مشخص محدود میکند.

به منظور کاهش نیاز شبکه به داده برچسب دار در این مقاله، بستر پیشنهادی برای بازسازی سه‌بعدی طوری طراحی شده است که بخش‌هایی از آن امکان این را دارند که از پیش به صورت بدون نظارت آموزش ببینند و در فاز آموزش با نظارت نیاز شبکه به داشتن برچسب برای آموزش از ابتدا[2] کاهش یابد. بنابراین بستر پیشنهادی نوعی روند نیمه نظارتی را برای یافتن راه حل در پیش خواهد گرفت. چون در آن هم از داده های برچب دار و هم بدون برچسب برای آموزش استفاده می‌شود.

ساختار ادامه مقاله به صورت زیر است: در بخش ۲ شامل مرور برخی کارهای مرتبط با بستر پیشنهادی است. ساختار شبکه پیشنهادی در این مقاله در بخش ۳ ارائه می‌شود و بخش ۴ شامل نتایج عددی و بصریِ مربوط به ارزیابی کارایی و مقایسه بستر پیشنهادی با برخی روش‌های موجود روی پایگاه‌های داده مختلف می‌باشد. نتیجه‌گیری و کارهای آینده در بخش ۵ بیان می‌شوند.

## ۲- مرور برخی تحقیقات مرتبط

در این بخش چند بستر یادگیری عمیق برای بازسازی سه‌بعدی از یک تصویر دوبعدی ورودی به عنوان کارهای مرتبط مرور و بررسی می‌شوند.

طبق تحقیق انجام شده در [۳]، شبکه‌های عمیق برای بازسازی سه‌بعدی از یک تصویر دوبعدی در واقع به جای عمل بازسازی، عملیات بازشناسی[3] انجام می‌شود. به این معنی که در این بسترها یک کتابخانه از اشکال سه‌بعدی آموخته می‌شود و در فاز آزمایش برای هر تصویر ورودی شکل متناظر موجود در کتابخانه آموخته شده به عنوان خروجی داده می‌شود.

در [۲]، یک شبکه عمیق پیشنهاد شده که در آنبه طور ضمنی سطح سه‌بعدی به صورت یک مرز تصمیم پیوسته در یک طبقه بند[4] یادگیری عمیق بیان می‌شود. این نحوه بیان شکل سه‌بعدی سه‌بعدی باعث می‌شود بتوانیم با اشکال سه‌بعدی با هر وضوحی[5] وضوحی[5] بدون مصرف حافظه زیاد کار کنیم. در این مقاله ما با استفاده از یک مدل سه‌بعدی به طور مستقیم با اشکال سه‌بعدی سروکار نداریم و بدین ترتیب مصرف حافظه بستر پیشنهادی در این مقاله اندک است.

بستر پیشنهادی را می‌توان با روش‌های بازسازی سه‌بعدی مبتنی بر مدل[6] مقایسه کرد [۱] که در آن‌ها با استفاده از روش‌هایی مانند[7](PCA) زیرفضاهایی با بعد پایین‌تر از داده مورد استفاده یافته می‌شود و این بازنمایی‌ها برای نگاشت مورد استفاده قرار می‌گیرند. در روش‌های کلاسیک مبتنی بر مدل زیرفضاهای یافته شده با استفاده از داده های محدودی یافته می‌شود به دلیل اینکه استفاده از روش PCA با داده زیاد پیچیدگی محاسباتی زیادی دارد. در بستر عمیق پیشنهادی در واقع عملکرد خود رمزگذارها شبیه روش PCA غیر خطی است چون در آنها نیز زیرفضایی با ابعاد پایین برای داده های مورد نظر یافته می‌شود. تفاوت خودرمز نگارهای عمیق در بستر پیشنهادی و روش PCA در این است که خود رمزگذارها محدودیت داده‌ی PCA را ندارند و هرچه تعداد داده ها برای آموزش این ساختارها بیشتر باشد زیرفضای مناسبتری بدون افزایش پیچیدگی محاسباتی یافته می‌شود.

در [۱۲] یک بستر یادگیری عمیق مبتنی بر تکرار[8] برای بازسازی سه‌بعدی چهره انسان از یک تصویر دوبعدی پیشنهاد شده است. در ساختارهای مبتنی بر تکرار در حین آموزش شبکه، خروجی شبکه به عنوان دانش جدید در مرحله بعد با داده ورودی همراه می‌شود و به بهبود کیفیت خروجی کمک می‌کند. این بستر از یک مدل سه‌بعدی به نام (3DMM)[9] برای تولید داده داده و برچسب برای آموزش استفاده می‌کند. همچنین یک فاز پس‌پردازش[10] با روش (SFS)[11] برای بهبود نتایج استفاده می‌شود. در تحقیق بعدی در [۱۳] نویسندگان همین مقاله با جایگزینی بخش SFS با یک شبکه عمیق دیگر کار قبلی خود را بهبود بخشیدند.

در [۱۶] یک بستر با نظارت برای بازسازی سه‌بعدی چهره انسان از یک تصویر ورودی پیشنهاد شده است که داده آموزش آن با استفاده از یک 3DMM بدست آمده است و در آن یک تابع

[1] Generative
[2] From scratch
[3] Recognition
[4] Classifier

[5] Resolution
[6] Model-based
[7] Principal Component Analysis
[8] Iterative
[9] ۳ D Morphable Model
[10] Post-process
[11] Shape From Shading

هزینه[1] جدید ارائه شده است که بین کلی بودن و فرامعین[2] بودن بودن شکل بدست آمده تعادل ایجاد می‌کند.

در بسترهای با نظارت پیشنهاد شده در [۱۲, ۱۳, ۱۶] علاوه بر حجم محاسباتی مورد نیاز برای تکراری بودن پردازش، نیاز به تعداد زیاد داده با برچسب و متعاقباً استفاده از مدل سه‌بعدی برای تولید داده قدرت این روش را به مدل سه‌بعدی به کار رفته برای تولید داده محدود می‌کند.

در [۱۱] یک بستر بدون نظارت برای بازسازی سه‌بعدی اشیاع از یک یا چند تصویر یا نقشه عمق[3] پیشنهاد شده است که در آن شکل خروجی با یک بخشِ ارائه دهنده به یک تصویر دو بعدی تبدیل می‌شود و با تصویر ورودی در فاز آموزش مقایسه می‌شود.

در [۱۴] یک ساختار شامل رمزگذار[4] و رمزگشا[5] برای آموزش بدون نظارت یک بستر برا بازسازی سه‌بعدی چهره انسان از یک تصویر دوبعدی پیشنهاد شده است. در این بستر در بخش بازنمایی رمزگذاری شده، شکل سه‌بعدی با اتخاذ فرض‌های مبتنی بر یک مدل سه‌بعدی، تشکیل میشود و در رمزگذار که یک ارائه دهنده ثابت است دوباره به تصویر دو بعدی تبدیل می‌شود.

بسترهای بدون نظارت نیز قدرت مدل را به قدرت ارائه دهنده مورد استفاده مدل سه‌بعدی مورد استفاده محدود می‌کند. بنابراین اگر بتوان در یک بستر، هم از داده واقعی به تعداد زیاد بدون برچسب و هم از داده برچسب دار بهره برد، می‌توان قدرت مدل طراحی شده را بالاتر برد.

در [۸] از یک ساختار (GAN)[6] برای تولید اشکال سه‌بعدی پیشنهاد شده که می‌تواند به عنوان تولید کننده داده مصنوعی در آموزش شبکه به کار رود.

در [۲۵] از یک شبکه پیچشی (CNN)[7] برای یادگیری با نظارت نظارت نگاشت تصویر دو بعدی به شکل سه‌بعدی استفاده شده است.

گرچه روش‌های موجود بر پایه یادگیری عمیق برای بازسازی سه‌بعدی به پیشرفت چشمگیری در بهبود کیفیت این حوزه دست یافته اند، اعتقاد ما براین است که هنوز می‌توان با اتخاذ فرض‌های مفید علاوه بر دستیابی به ساختار ساده تر بهبود کیفیت بازسازی را نیز به دست اورد. جزییات مربوط به بستر پیشنهادی در بخش بعد توضیح داده شده اند.

## ۳- بستر پیشنهادی

ایده طراحی بستر پیشنهادی در این مقاله، شکل دادن به بستری است که بتواند از اطلاعات برچسب در کنار داده‌های بدون برچسب از دنیای واقعی بهره بگیرد. براین مبنا از بخش های بدون نظارت در کنار بخش با نظارت بهره گرفته شده است که علاوه بر استفاده از اطلاعات مختلف، به شکل دهی به بستری با پارامتر کمتر می‌انجامد.

توضیح بستر پیشنهادی را با بیان ریاضی از حالت خاصی از فرایندی که در صدد انجام آن هستیم آغاز می‌کنیم. فرض کنید X نشان‌دهنده ماتریسی باشد که از کنار هم قرار دادن $n_1$ نمونه از تصاویر دو بعدی ($x_i$) که به صورت بردار یک بعدی مرتب شده اند، باشد.

$$X = [x_1, x_2, \dots, x_{n1}]_{D\times n_1} \quad (۱)$$

پس از حذف میانگین از X و اعمال Singular Value Decomposition (SVD) روی آن، پایه‌های زیرفضای PCA با رابطه زیر بدست می‌آید:

$$[U_x, \Sigma_{\mathrm{x}}, \mathrm{V_x}] = SVD(X) \quad (۲)$$

که در آن $U_x$ ماتریسی شامل بردارویژه‌های $XX^T$، $\Sigma_{\mathrm{x}}$ ماتریسی شامل مقدارویژه‌های $XX^T$ روی قطر اصلی آن و $\mathrm{V_x}$ نشان‌دهنده ماتریسی که ستون‌های آن بردارویژه‌های $X^TX$ می‌باشند. براین اساس می‌توانیم از ستون‌های ماتریس $U_x$ به عنوار بردارپایه‌های فضای بازنمایی برای تصاویر دوبعدی، استفاده کنیم. بنابراین بازنمایی k-بعدی حاصل برای تصاویر دوبعدی به صورت زیر بدست می‌آید:

$$Y_{k\times n_1} = U_x^T X_{D\times n_1} \quad (۳)$$

به طور مشابه با در نظر گرفتن Z به عنوان ماتریس شامل $n_2$ شکل سه‌بعدی ($z_i$) که به صورت بردار نمایش داده می‌شوند، داریم:

$$Z = [z_1, z_2, \dots, z_{n2}] \quad (۴)$$

$$[U_z, \Sigma_z, \mathrm{V_z}] = SVD(Z) \quad (۵)$$

$$B_{k'\times p} = U_z^T Z_{k'\times p} \quad (۶)$$

که در آن $B_{k'\times p}$ بازنمایی $k'$ - بعدی برای اشکال سه‌بعدی است. روابط (۳)و (۶) مراحل یافتن بازنمایی های بعدپایین برای تصاویر دوبعدی و اشکال سه‌بعدی در زیرفضاهای خطی هستند. ایده این مقاله یافتن نگاشت بین بازنمایی های بعد پایین است. در حالت

[1] Loss function
[2] Over-determined
[3] Depth map
[4] Encoder
[5] Decoder
[6] Generative Adversarial Network
[7] Convolutional Neural Network

خطی با روش کمترین مربعات خطا ماتریس نگاشت $P_{k\times k'}$ را بدست می‌آوریم:

$$P_{k\times k'} = argmin_T|B - TY| \quad (۷)$$

پس از یافتن ماتریس $T$ می‌توانیم شکل سه‌بعدی را از رابطه زیر بدست بیاوریم:

$$\hat{z} = U_z P U_x^T X \quad (۸)$$

ایده اصلی طراحی بستر پیشنهادی فراهم کردن یک سامانه غیر خطی و قدرمند است که حالت ساده ای از آن را با اتخاذ فرض‌های متعدد از قبیل خطی بودن زیرفضاها و عملگر نگاشت، در رابطه (۸) به طور شکل بسته بدست آوردیم. بر این مبنا بستر پیشنهادی در این مقاله از سه جزء اصلی ساخته شده است:

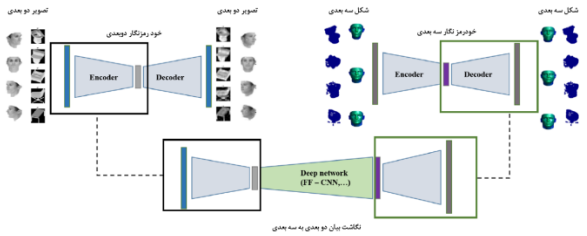
- خود رمزگذار دوبعدی[1] (2D AE) که به صورت بدون نظارت یک بازنمایی بعد پایین به طور غیر خطی برای تصاویر ورودی بدست می آورد.
- خود رمزگذار سه‌بعدی (3D AE) که به صورت بدون نظارت یک بازنمایی بعد پایین برای اشکال سه‌بعدی بدست می آورد.
- بخش نگاشت بازنمایی های دو بعدی به سه‌بعدی.

بنابر موارد ذکر شده، در بخش خود رمزگذار دوبعدی تصویر دوبعدی به بازنمایی بعدپایین نگاشت شده و در بخش نگاشت به بازنمایی بعد پایین سه‌بعدی نگاشت می‌شود و در نهایت بازنمایی بعد پایین سه‌بعدی در خود رمزگذار سه‌بعدی به شکل سه‌بعدی متناظر تبدیل می‌شود.

شکل (۱) دیاگرام بستر پیشنهادی برای بازسازی سه‌بعدی را نشان می‌دهد. که ما در آزمایشات خود برای بازسازی سه‌بعدی چهره انسان و اشیاء از یک تصویر دو بعدی از آن استفاده می‌کنیم.

جزئیات مربوط به ساختار هر بخش از بستر پیشنهادی در بخش ۴-۳ گزارش شده است.

آموزش اولیه خود رمزگذارها برای یافتن بازنمایی های معنی دارِ بعد پایین از تصاویر دو بعدی و اشکال سه‌بعدی بدون برچسب، باعث می‌شود بخش نگاشت کار نسبتاً راحتی را برای تبدیل بازنمایی دو بعدی به سه‌بعدی در پیش داشته باشد و بنابراین به پارامتر کمتر و در نتیجه داده برچسب دار کمتری برای آموزش دیدن نیاز دارد. اطلاعات مربوط به تعداد پارامترها و ساختار شبکه ها در بخش ۴-۳ گزارش داده شده است که طبق آن برای هر جزء بستر اعم از رمزگذارها و نگاشت حدود پنج لایه در نظر گرفته شده است که جمعا شبکه کوچکی برای بازسازی سه‌بعدی در مقایسه با شبکه های عمیق معادل [۱۲, ۱۳, ۱۶] تشکیل می‌دهد. این فرضیه را در بخش آزمایشات مورد بررسی قرار می‌دهیم.




شکل(۱) : نمودار بستر نیمه نظارتی پیشنهادی که در آن از دو خود رمزگذار برای یافتن بازنمایی های بعد پایین از تصاویر دو بعدی و اشکال سه‌بعدی به صورت بدون نظارت بدست می آید و سپس بخش نگاشت بازنمایی دو بعدی به سه‌بعدی بازنمایی بدست آمده دو بعدی را به بازنمایی سه‌بعدی تبدیل می‌کند.

## ۴- آزمایشات و شبیه سازی

در این بخش به منظور ارزیابی بستر پیشنهادی آزمایشات مختلفی را با استفاده از پایگاه‌های داده مختلف از چهره انسان برای بازسازی سه‌بعدی از یک تصویر ورودی انجام می‌دهیم.

### ۴-۱- پایگاه داده

به عنوان داده مورد استفاده در این مقاله برای چهره انسان هم از داده های دو بعدی برچسب دار با شکل سه‌بعدی متناظر استفاده شده است و هم از داده های بدون برچسب برای آموزش خود رمزگذارهای دو بعدی و سه‌بعدی. داده های برچسب دار از نوع مصنوعی انتخاب شده اند مانند پایگاه (BFM)[2] [۲۷] ، و برای آموزش خود رمزگذارها از داده های بدون برچسب دو بعدی و سه‌بعدی مانند LFW [۲۸] استفاده شده است.

همچنین از پایگاه داده Bosphorus [۲۹] که شامل داده های واقعی چهره انسان با برچسب سه‌بعدی متناظر هست اما تعداد داده ها در آن کم است برای تنظیم بیشتر[3] کل بستر استفاده شده است.

اندازه تصاویر دو بعدی ارائه[4] شده از پایگاه داده BFM برابر با $32 \times 32$ و $3 \times 227 \times 227$ در نظر گرفته شده است. تعداد ۴۰۰۰ نمونه داده آموزش و ۱۰۰۰ داده آزمایش از چهره انسان با حالت طبیعی چهره و از مقابل از این مدل سه‌بعدی تولید و مورد استفاده قرار گرفته است.

[1] 2D Auto Encoder (AE)

[2] Besel Face Model
[3] Fine tune
[4] Render

لازم به ذکر است که در مورد چهره انسان به دلیل نیاز به وضوح[1] زیاد در خروجی و به دلیل محدود بودن منابع محاسباتی و حافظه در اختیار، خروجی شبکه‌ها مستقیماً شکل سه‌بعدی نیستند و در واقع پارامترهای مدل سه‌بعدی هستند که شکل سه‌بعدی را نتیجه می‌دهد. پارامترهای مدل سه‌بعدی 3DMM ِBFM مورد استفاده بردارهایی با اندازه ۱۹۹ هستند که وقتی در بردار پایه های دخیره شده در مدل ضرب و حاصل با هم جمع می‌شود یک شکل سه‌بعدی را نتیجه می‌دهد. بنابرا این خروجی بستر های طراحی شده در آزمایشات $1 \times 199$ می‌باشد.

## ۴-۲- معیار ارزیابی

در این مقاله از مجموع مربعات خطا (MSE) به عنوان تابع هزینه برای آموزش شبکه ها و همچنین به عنوان معیار ارزیابی اشکال سه‌بعدی بدست آمده با هرروش استفاده می‌کنیم:

$$MSE = \sum_{i=1}^{N} \frac{|\hat{x}_o^i - \hat{x}_{GT}^i|}{N} \quad (۱)$$

که در آن $N$ تعداد داده ها در یک مجموعه آموزشی[2]، $\hat{x}_o^i$ خروجی شبکه و $\hat{x}_{GT}^i$ شکل هدف می‌باشد. این معیار میانگین اختلاف اشکال بدست آمده و واقعی را در یک مجموعه آموزشی از شبکه عصبی محاسبه می‌کند.

## ۴-۳- تنظیمات پارامترها

به منظور وجود تصاویر و اشکال سه‌بعدی با اندازه‌های مختلف در پایگاه‌های داده موجود در این مقاله پیکربندی‌های مختلفی برای بخش های بستر پیشنهادی طراحی شده است.

در مورد شکل سه‌بعدی چهره انسان که با بردارهای 199x1 سر وکار داریم، از لایه‌های dense برای طراحی رمزگذار سه‌بعدی استفاده کردیم. که اندازه لایه ها در آن به صورت 199-100-20-100-199 می باشد. همچنین بین لایه های موجود ازِتابع حالت LRelU و Batch normalization نیز استفاده شده است.

در مورد تصاویر دو بعدی با اندازه 32x32 از ساختار CNN برای طراحی رمزگذار استفاده کردیم که خروجی لایه‌ها در آن به صورت: 32x32x1-16x16x8-8x8x16-16x16x8-32x32x1 به همراه LRelU و Batch normalization و Maxpooling یا Upsampling پس از هر لایه می‌باشد.

در مورد تصاویر دوبعدی با اندازه 227x227x3 در بخش رمزگذار از شبکه آماده Alexnet [16] استفاده کردیم.

بخش نگاشت دوبعدی به سه‌بعدی در حالت چهره انسان از لایه‌های Dense به صورت 1024-500-200-100-20 به همراه LRelU و Batch normalization طراحی شده است و در حالت اشیاء به صورت 1024-1500-2000 به همراه LRelU و Batch normalization پس از هرلایه طراحی شده است.

خروجی بخش نگاشت برای تطابق با CNN رمزگشا[3] تغییر اندازه داده می‌شود.

بسترهای طراحی شده روی یک پردازنده گرافیکی با مدل NVIDIA GeForce GPU 1080 Ti با بهینه ساز ADAM آموزش دیدند.

## ۴-۴- نتایج عددی و بصری

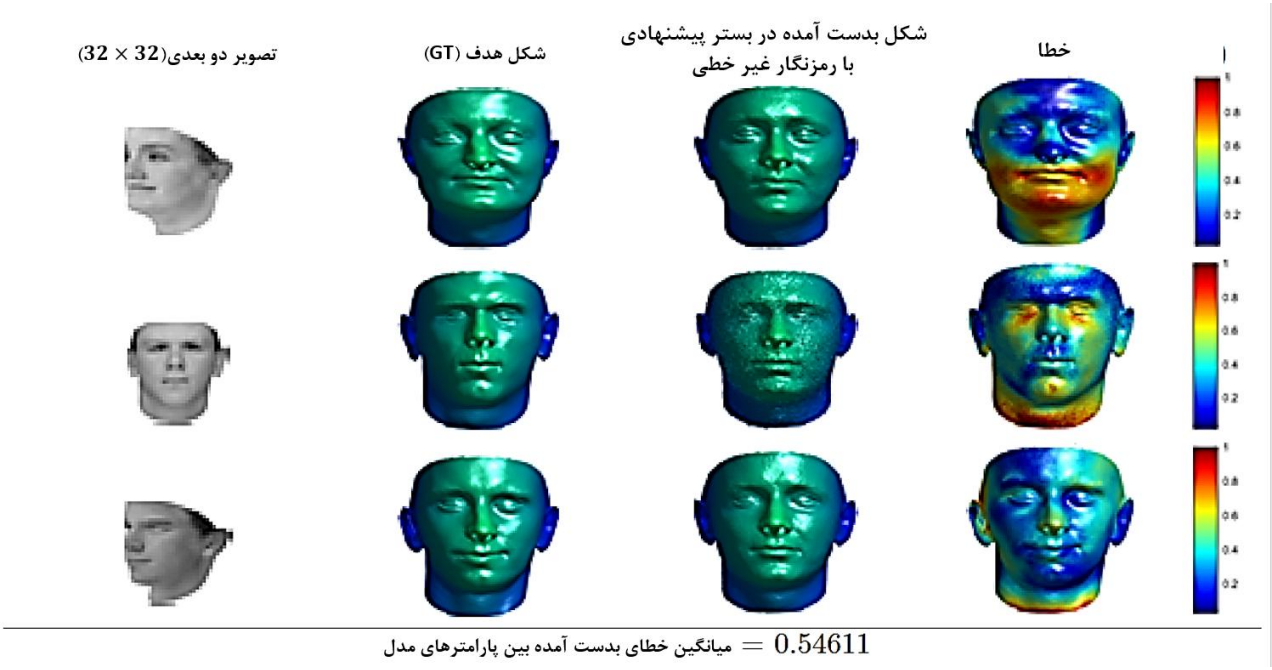
در این بخش نتایج عددی و بصری مربوط به بازسازی سه‌بعدی چهره انسان گزارش شده است.

## ۴-۵- بازسازی سه‌بعدی چهره انسان

برای بازسازی سه‌بعدی چهره انسان نیز ابتدا استفاده از رمزگذار غیرخطی را با حالتی که از شبکه Alexnet به عنوان رمزگذار استفاده می‌شود مقایسه می‌کنیم. شکل (۲) نشان دهنده نتایج بصری و عددی بازسازی سه‌بعدی با استفاده از رمزگذار غیرخطی است.

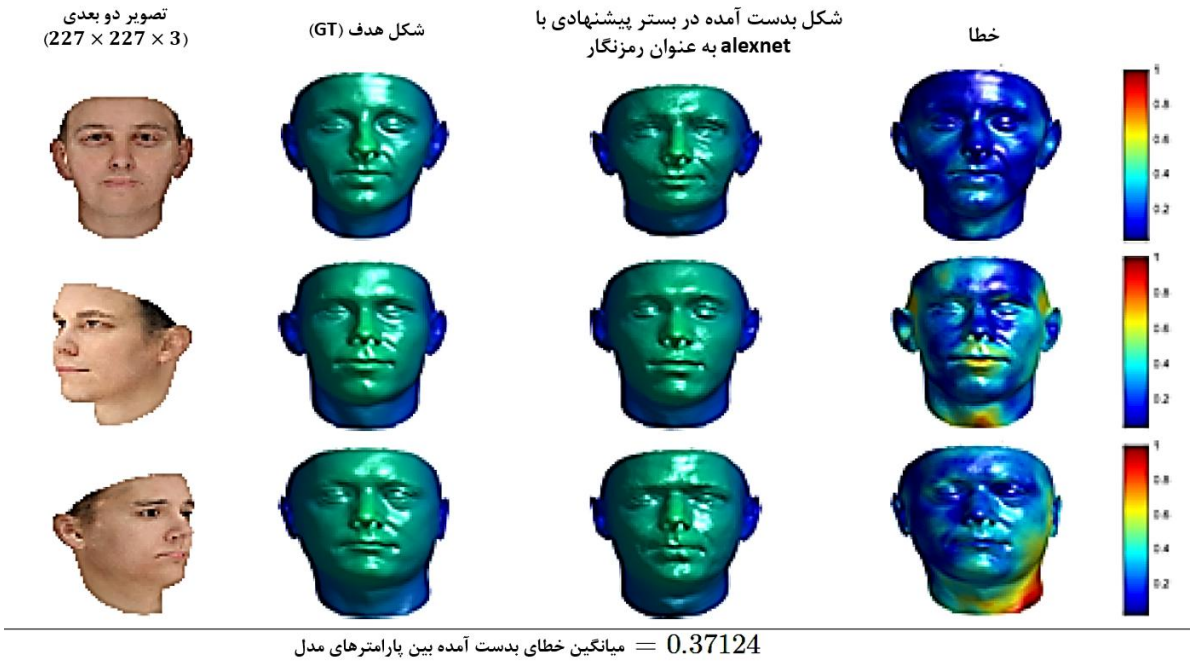
توجه به این نکته ضروری است که در حالت استفاده از پایگاه داده چهره انسان مستقیما با شکل سه‌بعدی کار نمی‌کنیم و خروجی شبکه بردار ۱۹۹ تایی از پارامترهای مدل سه‌بعدی به‌کاررفته BFM است. بنابراین خطای گزارش شده بین پارامتر بدست آمده توسط شبکه پیشنهادی و پارامتر مربوط به شکل سه‌بعدی هدف است. برای معنی دار شدن خطاهای گزارش شده و اینکه آیا بهبودی نسبت به حالت تصادفی بدست می‌دهند یا خیرریال آزمایشی انجام دادیم که در آن خطای بین شکل هدف و اشکال بدست آمده با پارامترهای تصادفی را بدست آوردیم. میانگین خطای بدست آمده بین پارامترهای تصادفی و پارامترهای شکل های هدف در ۲۰ اجرا برابر با ۰.۷۸۱ بدست آمد. بنابراین و با توجه به خطاهای گزارش شده در شکل (۴) و (۵) نشان می‌دهند که روش های پیشنهادی بهبودی هوشمند را نسبت به حالت تصادفی بدست می‌دهند.

[1] Resolution
[2] batch
[3] Decoder




شکل (۲): نتایج بصری و عددی بدست آمده از بکار بردن رمزگذار غیر خطی در بازسازی سه‌بعدی چهره انسان در بستر پیشنهادی

با جایگزین کردن رمزگذار غیر خطی با شبکه Alexnet نتایج موجود در شکل (۳) بدست می‌آیند.




شکل (۳): نتایج بصری و عددی بدست آمده از بکار بردن Alexnet به عنوان رمزگذاردر بازسازی سه‌بعدی چهره انسان در بستر پیشنهادی

نتایج گزارش شده در شکل‌های (۲) و (۳) نیز نشان می‌دهند باوجود بهبود خطای شبکه Alexnet نسبت به حالت استفاده از رمزگذار غیر خطی، میزان اختلاف خطا می‌تواند با نوجه به منبع محاسباتی خطا چشم پوشی شود. این به این دلیل است که نتایج بصری نشان داده شده در شکل (۲) نشان می‌دهد در حالت رمزگذار غیر خطی نیز نتایج مطلوبی بدست می‌آید.

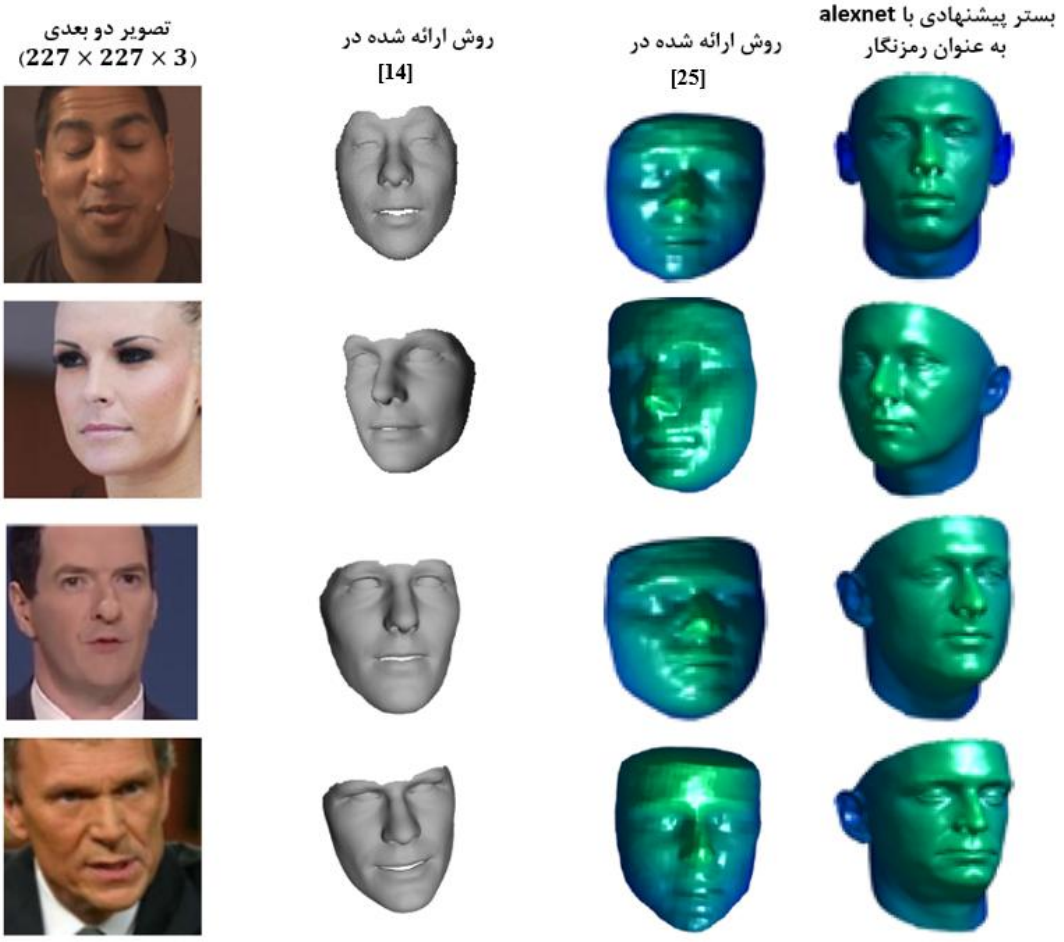
### ۴-۵-۱- مقایسه بصری با روش‌های دیگر

در این بخش کارایی بستر نیمه نظارتی پیشنهادی برای بازسازی سه‌بعدی از تصاویر واقعی مورد بررسی قرار می‌گیرد. به این منظور با استفاده از پایگاه داده چهره انسان Bosphorus که حاوی تصاویر واقعی از چهره ها است، بستر پیشنهادی با استفاده از Alexnet به عنوان رمزگذار و آموزش دیده با پایگاه BFM را تنطیم بیشتر کردیم و آن را برای بازسازی سه‌بعدی از تصاویر واقعی چهره های مختلف مورد استفاده قرار دادیم. در این مرحله با استفاده از تصاویری که توسط [۱۴] برای مقایسه در دسترس قرار داده شده به عنوان ورودی به شبکه استفاده نمودیم. هدف این بخش مقایسه بصری بستر پیشنهادی با دو روش ارائه شده در [۱۴, ۱۱] است.

نتایج بدست آمده توسط روش‌های موجود در شکل (۴) روی نمونه تصاویری که روش [۱۴] برای مقایسه در اختیار گذاشته است، نشان داده شده اند، که بیانگر این است بستر پیشنهادی به طور بصری توانسته است اشکال مطلوبی را برای هر تصویر ورودی نمونه بدست دهد.

## ۵- نتیجه

در این مقاله یک بستر نیمه نظارتی برای بازسازی سه‌بعدی چهره انسان و اشیاء از یک تصویر دو بعدی پیشنهاد شده است که از سه جزء اصلی خودرمزگذارها و بخش نگاشت تشکیل شده است. ایده اصلی در این مقاله یافتن نگاشتی بین بازنمایی‌های با ابعاد پایین تر تصاویر دو بعدی و اشکال سه‌بعدی است. نتایج عددی و بصری بدست آمده روی پایگاه های داده چهره انسان و اشیاء نشان می‌دهند که با وجود تعداد پارامتر کم به کار رفته در بستر پیشنهادی ، ساختار تفسیرپذیر آن به نتایج قابل قبولی نسبت به بسترهای پیچیده موجود برای بازسازی سه‌بعدی دست یافته است.




شکل (۴): نتایج بصری مقایسه بستر پیشنهادی با استفاده از شبکه Alexnet به عنوان رمزگذار با دو روش مرتبط بازسازی سه‌بعدی چهره انسان.

## مراجع


[1] Aldrian O, Smith WA. Inverse rendering of faces with a 3D morphable model. IEEE transactions on pattern analysis and machine intelligence. 2012 Sep 26;35(5):1080-93.

[2] Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S. and Geiger, A., 2019. Occupancy networks: Learning 3d reconstruction in function space. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4460-4470).

[3] Tatarchenko, M., Richter, S.R., Ranftl, R., Li, Z., Koltun, V. and Brox, T., 2019. What do single-view 3d reconstruction networks learn?. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3405-3414).

[4] Xu, Q., Wang, W., Ceylan, D., Mech, R. and Neumann, U., 2019. Disn: Deep implicit surface network for high-quality single-view 3d reconstruction. In Advances in Neural Information Processing Systems (pp. 492-502).

[5] Xie, H., Yao, H., Sun, X., Zhou, S. and Zhang, S., 2019. Pix2vox: Context-aware 3d reconstruction from single and multi-view images. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2690-2698).

[6] Fouhey DF, Gupta A, Zisserman A. Understanding higherorder shape via 3d shape attributes. IEEE TPAMI. 2017.

[7] Garrido P, Zollhöfer M, Casas D, Valgaerts L, Varanasi K, Pérez P, Theobalt C. Reconstruction of personalized 3D face rigs from monocular video. ACM Transactions on Graphics (TOG). 2016 May 18;35(3):1-5.

[8] Jiang L, Zhang J, Deng B, Li H, Liu L. 3d face reconstruction with geometry details from a single image. IEEE Transactions on Image Processing. 2018 Jun 8;27(10):4756-70.

[9] Kim K, Torii A, Okutomi M. Multi-view inverse rendering under arbitrary illumination and albedo. InEuropean conference on computer vision 2016 Oct 8 (pp. 750-767). Springer, Cham.

[10] Patow G, Pueyo X. A survey of inverse rendering problems. InComputer graphics forum 2003 Dec (Vol. 22, No. 4, pp. 663-687). Oxford, UK and Boston, USA: Blackwell Publishing, Inc.

[11] Rezende DJ, Eslami SA, Mohamed S, Battaglia P, Jaderberg M, Heess N. Unsupervised learning of 3d structure from images. InAdvances in neural information processing systems 2016 (pp. 4996-5004).

[12] Richardson E, Sela M, Kimmel R. 3D face reconstruction by learning from synthetic data. In2016 Fourth International Conference on 3D Vision (3DV) 2016 Oct 25 (pp. 460-469). IEEE.

[13] Richardson E, Sela M, Or-El R, Kimmel R. Learning detailed face reconstruction from a single image. InProceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017 (pp. 1259-1268).

[14] Tewari A, Zollhofer M, Kim H, Garrido P, Bernard F, Perez P, Theobalt C. Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction. InProceedings of the IEEE International Conference on Computer Vision Workshops 2017 (pp. 1274-1283).

[15] Thies J, Zollhofer M, Stamminger M, Theobalt C, Nießner M. Face2face: Real-time face capture and reenactment of rgb videos. InProceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 2387-2395).

[16] Tuan Tran A, Hassner T, Masi I, Medioni G. Regressing robust and discriminative 3D morphable models with a very deep neural network. InProceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 5163-5172).

[17] Wang J, Che C, Galeotti J, Horvath S, Gorantla V, Stetten G. Ultrasound tracking using ProbeSight: Camera pose estimation relative to external anatomy by inverse rendering of a prior high-resolution 3D surface map. In2017 IEEE Winter Conference on Applications of Computer Vision (WACV) 2017 Mar 24 (pp. 825-833). IEEE.

[18] Wood E, Baltrušaitis T, Morency LP, Robinson P, Bulling A. A 3d morphable eye region model for gaze estimation. InEuropean Conference on Computer Vision 2016 Oct 8 (pp. 297-313). Springer, Cham.

[19] Zhu X, Lei Z, Liu X, Shi H, Li SZ. Face alignment across large poses: A 3d solution. InProceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 146-155).

[20] Krawczyk-StańDo, D. and Rudnicki, M., 2007. Regularization parameter selection in discrete ill-posed problems—the use of the U-curve. International Journal of Applied Mathematics and Computer Science, 17(2), pp.157-164.

[21] Wolpert DH, Macready WG. No free lunch theorems for optimization. IEEE transactions on evolutionary computation. 1997 Apr;1(1):67-82.

[22] Gao Y, Yuille AL. Exploiting symmetry and/or Manhattan properties for 3D object structure estimation from single and multiple images. InProceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017 (pp. 7408-7417).

[23] Lun Z, Gadelha M, Kalogerakis E, Maji S, Wang R. 3D shape reconstruction from sketches via multi-view convolutional networks. In2017 International Conference on 3D Vision (3DV) 2017 Oct 10 (pp. 67-77). IEEE.

[24] Piotraschke M, Blanz V. Automated 3d face reconstruction from multiple images using quality measures. InProceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016 (pp. 3418-3427).

[25] Jackson AS, Bulat A, Argyriou V, Tzimiropoulos G. Large pose 3D face reconstruction from a single image via direct volumetric CNN regression. InProceedings of the IEEE International Conference on Computer Vision 2017 (pp. 1031-1039).

[26] Liu J, Yu F, Funkhouser T. Interactive 3D modeling with a generative adversarial network. In2017 International Conference on 3D Vision (3DV) 2017 Oct 10 (pp. 126-134). IEEE.

[27] Paysan P, Knothe R, Amberg B, Romdhani S, Vetter T. A 3D face model for pose and illumination invariant face recognition. In2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance 2009 Sep 2 (pp. 296-301). Ieee.

[28] Huang GB, Mattar M, Berg T, Learned-Miller E. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments.

[29] Savran A, Alyüz N, Dibeklioğlu H, Çeliktutan O, Gökberk B, Sankur B, Akarun L. Bosphorus database for 3D face analysis. InEuropean Workshop on Biometrics and Identity Management 2008 May 7 (pp. 47-56). Springer, Berlin, Heidelberg.

[30] Kim H, Zollhöfer M, Tewari A, Thies J, Richardt C, Theobalt C. Inversefacenet: Deep single-shot inverse face rendering from a single image. arXiv preprint arXiv:1703.10956. 2017.