# Package 'causalimages'

February 22, 2023

**Title** causalimages: R Package for Causal Inference with Earth Observation, Biomedical, and Social Science Images

**Version** 0.1

**Authors** 'Connor Jerzak <connor.jerzak@gmail.com> [aut, cre], Fredrik Johansson <fredrik.johansson@chalmers.se> [aut], Adel Daoud <adel.daoud@liu.se> [aut]'

**Description** R Package for causal inference with earth observation, biomedical, and social science images and image sequences (i.e., videos)

**Depends** R (>= 3.3.3)

**License**
Creative Commons Attribution-Noncommercial-No Derivative Works 4.0, for academic use only.

**Encoding** UTF-8

**LazyData** true

**Maintainer** 'Connor Jerzak' <connor.jerzak@gmail.com>

**Imports** tensorflow

**RoxygenNote** 7.2.1

## R topics documented:

---

AnalyzeImageHeterogeneity
                    *Decompose treatment effect heterogeneity by image*

---

### Description

Implements the image heterogeneity decomposition analysis of Jerzak, Johansson, and Daoud (2023). Users input in treatment and outcome data, along with a function specifying how to load in images using keys referenced to each unit (since loading in all image data will usually not be possible due to memory limitations). This function by default performs estimation, constructs salience maps, and can optionally perform estimation for new areas outside the original study sites in a transportability analysis.

**Usage**

```
AnalyzeImageHeterogeneity(obsW, obsY, acquireImageFxn, kClust_est, ...)
```

**Arguments**

| | |
|---|---|
| obsW | A numeric vector where 0's correspond to control units and 1's to treated units. |
| obsY | A numeric vector containing observed outcomes. |
| X | (optional) A numeric matrix containing tabular information used if `orthogonalize` = T. |
| orthogonalize | (default = F) A Boolean specifying whether to perform the image decomposition after orthogonalizing with respect to tabular covariates specified in X. |
| imageKeys | (default = 1:length(obsY)) A vector of length length(obsY) specifying the unique image ID associated with each unit. `imageKeys` are fed into `acquireImageFxn` to call images into memory. |
| kClust_est | (default = 2L) Integer specifying the number of clusters used in estimation. |
| acquireImageFxn | A function specifying how to load images associated with `imageKeys` into memory. For example, if observation 3 has a value of "a34f" in `imageKeys`, `acquireImageFxn` should extract the image associated with the unique key "a34f". First argument should be image key values and second argument have be `training` (in case behavior in training/) |
| acquireImageRepFxn | (default = acquireImageFxn) Similar to `acquireImageFxn`; this is a function specifying how to load images representations associated with `imageKeys` into memory. This will only be specified if there is a different |
| acquireImageFxn_transportability | (default = acquireImageFxn) Similar to `acquireImageFxn`; this is a function specifying how to load images for an optional transportability analysis, with transportability keys specified as a matrix in `transportabilityMat`. |
| transportabilityMat | (optional) A matrix with a column named keys specifying keys to be used by |
| long, lat | (optional) Vectors specifying longitude and latitude coordinates for units. Used only for describing highest and lowest probability neighorhood units if specified. |
| conda_env | (default = NULL) A string specifying a conda environment wherein `tensorflow`, `tensorflow_probability`, and `gc` are installed. |
| figuresKey | (default = "") A string specifying an identifier that is appended to all figure names. |
| figuresPath | (default = "./") A string specifying file path for saved figures made in the analysis. |
| simMode | (default = F) Should the analysis be performed in comparison with ground truth from simulation? |
| plotResults | (default = T) Should analysis results be plotted? |
| nDepthHidden_conv | (default = 3L) Hidden depth of convolutional layer. |
| nDepthHidden_dense | (default = 0L) Hidden depth of dense layers. Default of 0L means a single projection layer is performed after the convolutional layer (i.e., no hidden layers are used). |

maxPoolSize        (default = 2L) Integer specifying the max pooling size used in the convolutional
                   layers.

strides            (default = 2L) Integer specifying the strides used in the convolutional layers.=

yDensity           (default = normal) Specifies the density for the outcome. Current options in-
                   clude normal and lognormal.

nMonte_variational
                   (default = 5L) An integer specifying how many Monte Carlo iterations to use in
                   the calculation of the expected likelihood in each training step.

nMonte_predictive
                   (default = 10L) An integer specifying how many Monte Carlo iterations to use
                   in the calculation of posterior means (e.g., mean cluster probabilities).

nMonte_salience
                   (default = 100L) An integer specifying how many Monte Carlo iterations to use
                   in the calculation of the salience maps (e.g., image gradients of expected cluster
                   probabilities).

batchSize          (default = 25L) Batch size used in SGD optimization.

kernelSize         (default = 5L) Dimensions used in convolution kernels.

nSGD               (default = 400L) Number of stochastic gradient descent (SGD) iterations.

nDenseWidth        (default = 32L) Width of dense projection layers post-convolutions.

reparameterizationType
                   (default = "Flipout") Either "Flipout", or "Reparameterization". Speci-
                   fies the estimator used in the Bayesian neural components. With "Flipout",
                   convolutions are performed via CPU; with '"Reparameterization", they are per-
                   formed by GPU if available.

doConvLowerDimProj
                   (default = T) Should we project the nFilters convolutional feature dimensions
                   down to nDimLowerDimConv to reduce the number of required parameters.

nDimLowerDimConv
                   (default = 3L) If doConvLowerDimProj = T, then, in each convolutional layer, we
                   project the nFilters feature dimensions down to nDimLowerDimConv to reduce
                   the number of parameters needed.

nFilters           (default = 32L) Integer specifying the number of convolutional filters used.

quiet              (default = F) Should we suppress information about progress?

## Value

A list consiting of

- clusterTaus_mean default
- clusterTaus_sd Estimated image effect cluster standard deviations.
- clusterProbs_mean. Estimated mean image effect cluster probabilities.
- clusterTaus_sd. Estimated image effect cluster probability standard deviations.
- clusterProbs_lowerConf. Estimated lower confidence for effect cluster probabilities.
- impliedATE. Implied ATE.
- individualTau_est. Estimated individual-level image-based treatment effects.
- transportabilityMat. Transportability matrix withestimated cluster information.
- plottedCoordinates. List containing coordinates plotted in salience maps.
- whichNA_dropped. A vector containing observations dropped due to missingness.

## References

- Connor T. Jerzak, Fredrik Johansson, Adel Daoud. Image-based Treatment Effect Heterogeneity. Forthcoming in *Proceedings of the Second Conference on Causal Learning and Reasoning (CLeaR), Proceedings of Machine Learning Research (PMLR)*, 2023.

## Examples

```
#set seed
set.seed(1)

#Geneate data
x <- rnorm(100)
```

---

SimulateImageSystem          *Simulate causal systems involving images*

---

## Description

This function generates simulated causal structures using images. It is currently under construction.

## Usage

```
SimulateImageSystem(...)
```

## Arguments

| | |
|---|---|
| dag | *(character string)* An input DAG specifying causal structure. This input should be of the form '`i->t,i->y,t->y,....`' Currently, only one node in a DAG can be an image (this should be labeled "`i`"). The non-image nodes can have arbitrary string labels. The image can be a confounder, effect moderator, effect mediator. If the image is to be used as a moderator, use the notation, `t-i>y`. |
| ... | *(optional)* In estimation mode, users input the data matrices associated with the non-image nodes of `DAG` and image node `i`. For example, if `x` is a DAG node, users must, in estimation mode, supply data to `x` in a form that can be coerced to a tensor. |
| treatment | *(character string, optional)* In estimation mode, users specify the treatment variable here. If `treatment` is specified, users must provide other data inputs to the DAG (see `...`). |
| image_pool | *(character string, optional)* The path to where analysis specific images are located. This can be specified both in simulation and estimation mode. If not specified, the simulation uses a pool of Landsat images from Nigeria. |
| analysis_level | *(character string, default is '`scene`')* Defines the unit of analysis used in the simulation framework. This is ignored in estimation mode, where the unit of analysis is inferred from the data dimensions. |
| control | *(list)* A list containing control parameters in the data generating process. |

**Value**

A list:

- In *simulation mode*, the function returns a list with as many elements as unique nodes in DAG. Each element represents the simulated data.

- In *estimation mode*,the function returns an estimated treatment effect with 95\

**References**

- Connor T. Jerzak, Fredrik Johansson, Adel Daoud. Image-based Treatment Effect Heterogeneity. Forthcoming in *Proceedings of the Second Conference on Causal Learning and Reasoning (CLeaR), Proceedings of Machine Learning Research (PMLR)*, 2023.

**Examples**

```
#set seed
set.seed(1)

# Simulation mode
#simulatedData <- causalimage('r->i, i->t, t->y, r->y')
#print(names(simulatedData))

# Estimation mode
#estimatedResults <- causalimage('r->i, i->t, t->y, r->y', y=y, r=r, y=y', treatment='t')
#print( estimatedResults )
```

# Index