Abstract—By adhering to the dictum, “No causation without manipulation (treatment, intervention)”, cause and effect data analysis represents changes in observed data in terms of changes in the causal factors. When causal factors are not amenable for active manipulation in the real world due to current technological limitations or ethical considerations, a counterfactual approach performs an intervention on the model of data formation. In the case of object representation or activity (temporal object) representation, varying object parts is generally unfeasible whether they be spatial and/or temporal. Multilinear algebra, the algebra of higher order tensors, is a suitable and transparent framework for disentangling the causal factors of data formation. Learning a part-based intrinsic causal factor representations in a multilinear framework requires applying a set of interventions on a part-based multilinear model. We propose a unified multilinear model of wholes and parts. We derive a hierarchical block multilinear factorization, the M-mode Block SVD, that computes a disentangled representation of the causal factors by optimizing simultaneously across the entire object hierarchy. Given computational efficiency considerations, we introduce an incremental bottom-up computational alternative, the Incremental M-mode Block SVD, that employs the lower level abstractions, the part representations, to represent the higher level of abstractions, the parent wholes. This incremental computational approach may also be employed to update the causal model parameters when data becomes available incrementally. The resulting object representation is an interpretable combinatorial choice of intrinsic causal factor representations related to an object’s recursive hierarchy of wholes and parts that renders object recognition robust to occlusion and reduces training data requirements.

Index Terms—causality, counterfactuals, explanatory variables, latent representation, factor analysis, tensor algebra, M-mode SVD, block tensor decomposition, hierarchical block tensor factorization, hierarchical tensor, structural equation model, object recognition, image analysis, data augmentation

I. INTRODUCTION: PROBLEM DEFINITION

Developing causal explanations for correct results or for failures from mathematical equations and data is important in developing a trustworthy artificial intelligence, and retaining public trust. Causal explanations are germane to the “right to an explanation” statute [15], [13] i.e., to data driven decisions, such as those that rely on images. Computer graphics and computer vision problems, also known as forward and inverse imaging problems, have been cast as causal inference questions [40], [42] consistent with Donald Rubin’s quantitative definition of causality, where “A causes B” means “the effect of A is B”, a measurable and experimentally repeatable quantity [14], [17]. Computer graphics may be viewed as addressing analogous questions to forward causal inferencing that addresses the “what if” question, and estimates the change in effects given a delta change in a causal factor. Computer vision may be viewed as addressing analogous questions to inverse causal inferencing that addresses the “why” question [12]. We define inverse causal inference as the estimation of causes given an estimated forward causal model and a set of observations that constrain the solution set.

Natural images are the compositional consequence of multiple factors related to scene structure, illumination conditions, and imaging conditions. Multilinear algebra, the algebra of higher-order tensors, offers a potent mathematical framework for analyzing the multifactor structure of image ensembles and for addressing the difficult problem of disentangling the constituent factors, Fig. 2. (Vasilescu and Terzopoulos: TensorFaces 2002 [33], [44], MPCA and MICA 2005 [46], kernel variants [40], Multilinear Projection 2007/2011[47], [41])

Fig. 1: Data tensor, $\mathcal{D}$, expressed in terms of a hierarchical data tensor, $\mathcal{D}_h$, a mathematical instantiation of a tree data structure where $\mathcal{D} = \mathcal{D}_h \times_1 I_1 \times_2 I_2 \times_3 I_3 \ldots \times_n I_n$, versus a bag of independent parts/sub-parts, or a data tensor with a reparameterized measurement mode in terms of regions and sub-regions. An object hierarchy may be based on adaptive quad/triangle based subdivision of various depths [35], or a set of perceptual parts of arbitrary shape, size and location. Images of non-articulated objects are best expressed with hierarchical data tensors that have a partially compositional form, where all the parts share the same extrinsic causal factor representations, Fig. 3b. Images of objects with articulated parts are best expressed in terms of hierarchical data tensors that are fully compositional in the causal factors, Fig. 3c. Images of non-articulated objects may also be represented by a fully compositional hierarchical data tensor, as depicted by the TensorTrinity example above.
Scene structure is composed from a set of objects that appear to be formed from a recursive hierarchy of perceptual wholes and parts whose properties, such as shape, reflectance, and color, constitute a hierarchy of intrinsic causal factors of object appearance. Object appearance is the compositional consequence of both an object’s intrinsic causal factors, and extrinsic causal factors with the latter related to illumination (i.e. the location and types of light sources), and imaging (i.e. viewing direction, camera lens, rendering style etc.). Intrinsic and extrinsic causal factors confound each other’s contributions hindering recognition [42].

“Intrinsic properties are by virtue of the thing itself and nothing else” (David Lewis, 1983 [22]); whereas extrinsic properties are not entirely about that thing, but as a result of the way the thing interacts with the world. Unlike global intrinsic properties, local intrinsic properties are intrinsic to a part of the thing, and it may be said that a local intrinsic property is in an "intrinsic fashion", or “intrinsically” about the thing, rather than “is intrinsic” to the thing [19].

Cause and effect analysis models the mechanisms of data formation, unlike conventional statistical analysis and conventional machine learning that model the distribution of the data [29]. Causal modeling from observational studies are suspect of bias and confounding with some exceptions [8], [34], unlike experimental studies [31], [32] in which a set of active interventions are applied, and their effect on response variables are measured and modeled. The differences between experimental studies, denoted symbolically with Judea Pearl’s do-operator [29], and observational studies are best exemplified by the following expectation and probability expressions

\[
E(d|c) \neq E(d(do(c)))
\]

From Observational Studies:

\[
P(d|c) \neq P(d(do(c)))
\]

From Experimental Studies:

Association, Correlation, Prediction Causation

where \(d\) is a multivariate observation, and \(c\) is a hypothesized or actual causal factor. Pearl and Bareinboim [30], [2] have delineated the challenges of generalizing results from experimental studies to observational studies by parameterizing the error based on the possible error inducing sources.

The multilinear (tensor) structural equation approach is a suitable view for viewing direction, camera lens, rendering style etc.). Intrinsic representation, varying object parts is generally unfeasible in the case of object representation or activity (temporal object) consequence of both an object’s intrinsic causal factors, and extrinsic causal factors confound each other’s contributions hindering recognition [42].

The multilinear (tensor) structural equation approach is a suitable representation, varying object parts is generally unfeasible whether they be spatial or temporal. Learning a hierarchy of intrinsic causal factor representations requires applying a set of interventions on the structural model, hence it requires a part-based multilinear model, Fig 1.

This paper proposes a unified multilinear model of wholes and parts that defines a data tensor in terms of a hierarchical data tensor, \(D_m\), a mathematical instantiation of a tree data structure. Our hierarchical data tensor is a mathematical conceptual device that allows for a different tree parameterization for each causal factor, and enables us to derive a multilinear hierarchical block factorization, an M-mode Block SVD, that optimizes simultaneously across the entire object hierarchy. Given computational considerations, we develop an incremental computational alternative that employs the lower level abstractions, the part representations, to represent the higher level of abstractions, the parent wholes.

Our hierarchical block multilinear factorization, M-mode Block SVD, disentangles the causal structure by computing statistically invariant intrinsic and extrinsic representations. The factorization learns a hierarchy of low-level, mid-level and high-level features. Our hybrid approach mitigates the shortcomings of local features that are sensitive to local deformations and noise, and the shortcomings of global features that are sensitive to occlusions. The resulting object representation is a combinatorial choice of part representations, that renders object recognition robust to occlusion and reduces large training data requirements. This approach was employed for face verification by computing a set of causal explanations (causalX) [42].

II. RELEVANT TENSOR ALGEBRA

We will use standard textbook notation, denoting scalars by lower case italic letters \((a, b, ...\)\), vectors by bold lower case letters \((\mathbf{a}, \mathbf{b}, ...\)\), matrices by bold uppercase letters \((\mathbf{A}, \mathbf{B}, ...\)\), and higher-order tensors by bold uppercase calligraphic letters \((\mathcal{A}, \mathcal{B}, ...\)\). Index upper bounds are denoted by italic uppercase \((i.e., 1 \leq i \leq I)\). The zero matrix is denoted by \(0\), and the identity matrix is denoted by \(I\). References [21], [33] provide a quick tutorial, but references [40], [46], [41] are an indepth treatment of tensor based factor analysis.

Briefly, the natural generalization of matrices (i.e., linear operators defined over a vector space), tensors define multilinear operators over a set of vector spaces. A “data tensor” denotes an \(M\)-way data array.

**Definition 1 (Tensor):** Tensors are multilinear mappings over a set of vector spaces, \(\mathbb{R}^{I_1} \times \cdots \times \mathbb{R}^{I_C}\), to a range vector space \(\mathbb{R}^{I_0}\):

\[
A : \{\mathbb{R}^{I_1} \times \cdots \times \mathbb{R}^{I_C}\} \rightarrow \mathbb{R}^{I_0}.
\]

The order of tensor \(A \in \mathbb{R}^{I_0 \times I_1 \times \cdots \times I_C}\) is \(M = C + 1\). An element of \(A\) is denoted as \(a_{i_1...i_c}\) or \(a_{0\cdots 0...i_c}\), where \(1 \leq i_0 \leq I_0\), and \(1 \leq i_c \leq I_c\).

The mode-\(m\) vectors of an \(M\)-order tensor \(A \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_M}\) are the \(I_m\)-dimensional vectors obtained from \(A\) by varying \(i_m\) while keeping the other indices fixed. In tensor terminology, column vectors are the mode-1 vectors and row vectors as mode-2 vectors. The mode-\(m\) vectors of a tensor are also known as fibers. The mode-\(m\) vectors are the column vectors of matrix \(A_{[m]}\) that results from matrixizing (a.k.a. flattening) the tensor \(A\).
Within the tensor mathematical framework, a generalization of the product of two matrices is the product of a tensor and a matrix \[ A \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_M} \]. As the parenthetical ordering indicates, the mode-\( m \) column vectors are arranged by sweeping all the other mode indices through their ranges, with smaller mode indexes varying more rapidly than larger ones; thus,

\[
[A_{[m]}]_{jk} = a_{i_1 \ldots i_m \ldots i_M}, \quad \text{where} \quad j = i_m \quad \text{and} \quad k = 1 + \sum_{n=0}^{M} (i_n - 1) \prod_{l=0 \atop l \neq m}^{n-1} I_l.
\]

A generalization of the product of two matrices is the product of a tensor and a matrix [9].

Definition 3 (Mode-\( m \) Product, \( \times_m \)): The mode-\( m \) product of a tensor \( A \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_M} \) and a matrix \( B \in \mathbb{R}^{I_m \times I_m} \), denoted by \( A \times_m B \), is a tensor of dimensionality \( \mathbb{R}^{I_1 \times \ldots \times I_{m-1} \times I_m \times I_{m+1} \times \ldots \times I_M} \) whose entries are computed by

\[
[A \times_m B]_{i_1 \ldots i_{m-1} i_m i_{m+1} \ldots i_M} = \sum_{i_m} a_{i_1 \ldots i_{m-1} i_m i_{m+1} \ldots i_M} b_{i_m i_m},
\]

\[
C = A \times_m B \quad \text{matrixize tensorize} \quad C_{[m]} = BA_{[m]}.
\]

The \( M \)-mode SVD (aka. the Tucker decomposition) is a “generalization” of the conventional matrix (i.e., 2-mode) SVD which may be written in tensor notation as

\[
D = U_1 S U_2^T \quad \Leftrightarrow \quad D = S \times_1 U_1 \times_2 U_2.
\]

The \( M \)-mode SVD orthogonalizes the \( M \) spaces and decomposes the tensor as the mode-\( m \) product, denoted \( \times_m \), of \( M \)-orthonormal mode matrices, and a core tensor \( D \)

\[
D = Z \times_1 U_1 \times_2 U_2 \cdots \times_m U_m \cdots \times_M U_M.
\]
Fig. 3: The data tensor, \( \mathcal{D} \), written in terms of a hierarchical data tensor, \( \mathcal{D}_H \). (a) When \( \mathcal{D}_H \) contains the data tensor segments, \( \mathcal{D}_s \), along its super-diagonal then \( \mathcal{D}_H \) has a \textit{fully compositional} form, and every mode matrix has a compositional representation. (b) A general base case object written in a \textit{partially compositional} form with a compositional representation for only one mode matrix (causal factor). (c) A general base case object where all the causal factors have a compositional representation. The tensor \( \mathcal{D}_N \) is \textit{fully compositional in the causal factors}. (d) A base case object with non-overlapping parts. All the causal factors have a compositional representations. Multilinear factorizations are block independent. (e) Base case object with completely overlapping parts where all the causal factors have compositional representation. Objects with non-overlapping or completely overlapping parts may also be written using a partially compositional form analogously to (b).

Multivariate array, \( \mathbf{D} \in \mathbb{R}^{I_{x_r} \times I_{x_c}} \), with \( I_{x_r} \) rows and \( I_{x_c} \) columns, the convolution is written as

\[
\mathbf{D}_s = \mathbf{D} \ast \mathbf{h}_s(x, y) \overset{\text{vectorsize}}{\longrightarrow} \mathbf{d} = \mathbf{H} \mathbf{d} = \mathbf{d} \times_0 \mathbf{H} \tag{6}
\]

where the measurement mode is mode 0. In practice, a convolution is efficiently implemented using a DFFT. The segment data tensor, \( \mathcal{D}_s = \mathcal{D} \times_0 \mathbf{H}_s \), is the result of multiplying (convolving) every observation, \( \mathbf{d} \), with the block circulant matrix (filter), \( \mathbf{H}_s \). A filter \( \mathbf{H}_s \) may be of any type, and have any spatial scope. When a filter matrix is a block identity matrix, \( \mathbf{H}_s = \mathbf{I} \), the filter matrix multiplication with a vectorized observation has the effect of segmenting a portion of the data. Measurements associated with perceptual parts may not be tightly packed into a block apriori, as in the case of vectorized images, but chunking is achieved by a trivial permutation.

A data tensor is expressed as a recursive hierarchy of wholes and parts by defining and employing a \textit{hierarchical data tensor}, \( \mathcal{D}_H \). When a data tensor contains along its super-diagonal the data tensor segments, \( \mathcal{D}_s \), then \( \mathcal{D}_H \) has a \textit{fully compositional} form, and all the data tensor modes have a compositional representation, Fig. 3(a). The data tensor segments, \( \mathcal{D}_s \), may be sparse and represent local parts, or may be full and correspond to a filtered version of a parent-whole, as in the case of a Laplacian pyramid. Mathematically writing \( \mathcal{D} \) in terms of \( \mathcal{D}_H \) is expressed with

\[
\mathcal{D}_s = \sum_{s=1}^{S} \mathcal{D} \times_s \mathbf{H}_s \tag{7}
\]

\[
\mathcal{D}_N = \mathcal{D}_s \cdots + \mathcal{D}_s \tag{8}
\]

\[
\mathcal{D}_N = \mathcal{D}_s \times_0 \mathbf{I}_{x_0} \times_1 \mathbf{I}_{x_1} \cdots \times_c \mathbf{I}_{x_c} \tag{9}
\]

where \( \mathbf{I}_{x_c} = [I_{x_1}, \ldots, I_{x_c}] \in \mathbb{R}^{I_{x_c} \times SI_c} \) is a concatenation of \( S \) identity matrices, one for each data segment. In practice, the measurement mode will not be written in compositional form, ie. the multipication with \( \mathbf{I}_{x_0} \), would have been carried out. The resulting \( \mathcal{D}_N \) is \textit{fully compositional in the causal factors}, where every causal factor has a compositional representation rather than every mode Fig. 3(c). Articulated-objects have parts with their own extrinsic causal factors and benefit from a compositional representation of every causal factor. A non-articulated object where the wholes, and parts share the same extrinsic causal factor representations (same illumination/viewing conditions) benefit from being written in terms of a \textit{partially compositional
data tensor, where the intrinsic causal factor has a compositional form, the intrinsic object representation, Fig. 3(b). Thus, the $D_{n}$ is multiplied through by all the $I_{c}$ except one. Each multiplied $I_{c}$ is replaced by a single place holder identity matrix in the model.

The three different ways of rewriting $D$ in terms of a hierarchy of wholes and parts, eq. 7-9, results in three mathematically equivalent representations based on factorizing $D$, $D_{j}$, and $D_{n}$:

$$D = \sum_{s=1}^{S} (Z_{s} \times U_{s \times \cdot \cdot \cdot \times U_{c}}) \times H_{c}$$

$$D = \sum_{s=1}^{S} (Z_{s} \times U_{s \times \cdot \cdot \cdot \times U_{c}}) \times H_{c}$$

$$D = \sum_{s=1}^{S} (Z_{s} \times U_{s \times \cdot \cdot \cdot \times U_{c}}) \times H_{c}$$

Despite the prior mathematical equivalence, equations 7,10, and equations 8,12 are not flexible enough to explicitly indicate if the parts are organized in a partially compositional form, or a fully compositional form.

The expression of $D$ in terms of a hierarchical data tensor is a mathematical conceptual device, that enables a unified mathematical model of wholes and parts that can be expressed completely as a mode-m product (tensor-matrix multiplication) and whose factorization can be optimized in a principled manner.

Dimensionality reduction of the compositional representation is performed by optimizing

$$e = \|D - (\bar{Z}_{n} \times \bar{U}_{n \times \cdot \cdot \cdot \times U_{c \times \cdot \cdot \cdot \times U_{c}}}) \times I_{c \times \cdot \cdot \cdot \times I_{c}}\|^{2}$$

where $U_{c \times \cdot \cdot \cdot \times U_{c}}$ is the composite representation of the $c$th mode, and $\bar{Z}_{n}$ governs the interaction between causal factors. Our optimization may be initialized by setting $\bar{Z}_{n}$ and $U_{n \times \cdot \cdot \cdot \times U_{c \times \cdot \cdot \cdot \times U_{c}}}$ to the M-mode SVD of $D_{n}$, 4, 5 and performing dimensionality reduction through truncation, where $U_{n \times \cdot \cdot \cdot \times U_{c \times \cdot \cdot \cdot \times U_{c}}}$ is in $\mathbb{R}_{0 \times \cdot \cdot \cdot \times I_{c \times \cdot \cdot \cdot \times I_{c}}}$ and $J_{c} \leq S_{I}$. 

B. Derivation

For notational simplicity, we re-write the loss function as,

$$e := \|D - \bar{Z}_{n} \times \bar{U}_{n \times \cdot \cdot \cdot \times U_{c \times \cdot \cdot \cdot \times U_{c}}}) \times I_{c \times \cdot \cdot \cdot \times I_{c}}\|^{2}$$

where $\bar{U}_{n \times \cdot \cdot \cdot \times U_{c \times \cdot \cdot \cdot \times U_{c}}})$ is permutation matrix that groups the columns of $U_{n \times \cdot \cdot \cdot \times U_{c \times \cdot \cdot \cdot \times U_{c}}}$ based on the segment, $s$, to which they belong, and the inverse permutation matrices have been multiplied into $\bar{Z}_{n}$ resulting into a core that has also been grouped based on segments and sorted based on variance. The data tensor, $D$, may be expressed in matrix form as in eq. 16 and reduces to the more efficiently structured as in eq. 17

$$D = \bar{Z}_{n} \times U_{n \times \cdot \cdot \cdot \times U_{c \times \cdot \cdot \cdot \times U_{c}}})$$

$$D = \bar{Z}_{n} \times U_{n \times \cdot \cdot \cdot \times U_{c \times \cdot \cdot \cdot \times U_{c}}})$$

$$D = \bar{Z}_{n} \times U_{n \times \cdot \cdot \cdot \times U_{c \times \cdot \cdot \cdot \times U_{c}}})$$

The Kronecker product of $U \in \mathbb{R}^{I \times J}$ and $V \in \mathbb{R}^{K \times L}$ is the $IK \times JL$ matrix defined as $[U \otimes V]_{i \times j \times k \times l} = u_{i \times j \times k \times l}$. 6

The Kronecker product of $U \in \mathbb{R}^{I \times J}$ and $V \in \mathbb{R}^{K \times L}$ is a block-matrix Kronecker product; therefore, it can be expressed as $U \otimes V = ([U_{1} \otimes V_{1}] \cdot \cdot \cdot [U_{S} \otimes V_{S}])$. 7

where $\otimes$ is the Kronecker product, and $\otimes$ is the block-matrix Kahtri-Rao product.

The matrixized block diagonal form of $Z_{n}$ in eq. 17 becomes evident when employing our modified data centric matrixizing operator based on the definition 2, where the initial mode is the measurement mode.

The hierarchical block multilinear factorization, the $M$-mode Block SVD algorithm computes the matrix mode, $U_{n}$, by computing the minimum of $e = \|D - \bar{Z}_{n} \times U_{n \times \cdot \cdot \cdot \times U_{c \times \cdot \cdot \cdot \times U_{c}}})\|^{2}$ by cycling through the modes, solving for $U_{n}$ in the equation $\partial e/\partial U_{n} = 0$ while holding the core tensor $\bar{Z}_{n}$ and all other mode matrices constant, and repeating until convergence. Note that

$$\frac{\partial e}{\partial U_{n}} = \frac{\partial}{\partial U_{n}} \|D - U_{n}W_{n}^{T}\|^{2} = -2D_{t}W_{n}^{T} + 2U_{n}W_{n}^{T}$$

Thus, $\partial e/\partial U_{n} = 0$ implies that

$$U_{n} = D_{t}(W_{n}^{T}W_{n})^{-1}D_{1}W_{n}^{T} = D_{t}(\tilde{Z}_{n})^{T}$$

$$U_{n} = D_{t}(W_{n}^{T}W_{n})^{-1}D_{1}W_{n}^{T} = D_{t}(\tilde{Z}_{n})^{T}$$

$$U_{n} = D_{t}(W_{n}^{T}W_{n})^{-1}D_{1}W_{n}^{T} = D_{t}(\tilde{Z}_{n})^{T}$$

$$U_{n} = D_{t}(W_{n}^{T}W_{n})^{-1}D_{1}W_{n}^{T} = D_{t}(\tilde{Z}_{n})^{T}$$

$$U_{n} = D_{t}(W_{n}^{T}W_{n})^{-1}D_{1}W_{n}^{T} = D_{t}(\tilde{Z}_{n})^{T}$$

whose $U_{n \times \cdot \cdot \cdot \times U_{c \times \cdot \cdot \cdot \times U_{c}}}$ sub-matrices are then subject to orthonormality constraints.

Solving for the optimal core tensor, $Z_{n}$, the data tensor, $D$, approximation is expressed in vector form as,

$$e = \|\text{vec}(D) - (U_{c} \otimes \cdot \cdot \cdot \otimes U_{c} \otimes \cdot \cdot \cdot \otimes U_{n})\|^{2}(\bar{Z}_{n})\|^{2}.$$
Algorithm 1 M-mode Block SVD.

Input: Data tensor, $\mathcal{D} \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_c}$, filters $H_i$, and desired dimensionality reduction $J_1, \ldots, J_c$.

1. Initialization:
   1a. Decompose each data tensor segment, $\mathcal{D}_i = \mathcal{D} \times H_i$, by employing the M-mode SVD.
   \[ \mathcal{D} = Z \times U_{i,0} \times \cdots \times U_{i,c} \times V_{i,0} \times \cdots \times V_{i,c} \]
   1b. For $\epsilon = 0, 1, \ldots, C$, set $U_{i,\epsilon} = [U_{i,c,0}, \ldots, U_{i,c,\epsilon}]$, and truncate to $J_i$ columns by sorting all the eigenvalues from $U_{i,c,0}$ and the rows from $Z_{i,c}$. 

2. Optimization via alternating least squares:
   Iterate for $n = 1, \ldots, N$
   For $\epsilon = 0, \ldots, C$
   2a. Compute mode matrix $U_{i,\epsilon}$ while holding the rest fixed.
   \[ U_{i,\epsilon} = D_{i,\epsilon} \left( U_{i,c,0} \otimes \cdots \otimes U_{i,c,\epsilon} \otimes \cdots \otimes D_{i,c} \right)^T \]
   until convergence.

Output converged matrices $U_{1,\epsilon}, \ldots, U_{c,\epsilon}$ and tensor $Z_{i,\epsilon}$.

When the data tensor is a collection of observations made up of non-overlapping parts, Fig. 3d, the data tensor decomposition reduces to the concatenation of an M-mode SVD of individual parts and when the data tensor is a collection of overlapping parts that have the same multilinear-rank reduction, Fig. 3e, see [42] for additional specific optimizations.

IV. REPRESENTING LEVELS OF ABSTRACTION BOTTOM-UP
An incremental hierarchical block multilinear factorization that represents levels of abstractions bottom-up is developed analogously to the incremental SVD for matrices [4]. The precomputing multilinear factorizations of the children parts are employed to determine the parent whole multilinear factorization. The derived algorithm may also be employed to update the overall model when the data becomes available sequentially [23]. We first address the computation of the mode matrices and the extended core of the parent whole when the children parts are non-overlapping. Next, we consider the overlapping children case, and the case where the parent-wholes and children-parts contain differently filtered data.

Computing parent causal mode matrices, $U_{c,\epsilon}$: Note that the parent whole, $\mathcal{D}_{c,\epsilon}$ is a concatenation of the data contained in its $K' \epsilon$ children segments that are part of the hierarchy, $\mathcal{D}_k$, where $1 \leq k \leq K'$. New data that is not contained by any of the children is denoted as the $K = K' + 1$ child, $\mathcal{D}_c$, eq. 21. We initialize the hierarchical block multilinear factorization by performing an $M$-mode SVD on each leaf.

The mode matrix, $U_{c,\epsilon}$ of the $\epsilon$ parent whole, $\mathcal{D}_{c,\epsilon}$, is the left singular matrix of \[ U_{c,\epsilon} \Sigma_{\epsilon,1} \cdots U_{c,\epsilon} \Sigma_{\epsilon,k} \cdots U_{c,\epsilon} \Sigma_{\epsilon,K} \] which is based on the following derivation, that writes SVD of the flattened parent whole in terms of the SVDs of its flattened children parts, followed by a collection terms such that $V_{c,\epsilon}$ is a block diagonal matrix of $V_{c,k}$:

\[ D_{c,\epsilon} = \begin{bmatrix} D_{c,1} & \cdots & D_{c,k} & \cdots & D_{c,K} \end{bmatrix} \]

\[ = \begin{bmatrix} U_{c,1} \Sigma_{c,1} T_{1i0}(U_{c,1} \otimes \cdots \otimes U_{c,\epsilon} \otimes \cdots \otimes U_{c,\epsilon} \otimes \cdots \otimes U_{c,K})^T \cdots U_{c,\epsilon} \Sigma_{c,k} T_{1i0}(U_{c,1} \otimes \cdots \otimes U_{c,\epsilon} \otimes \cdots \otimes U_{c,\epsilon} \otimes \cdots \otimes U_{c,K})^T \cdots \end{bmatrix} \]

\[ = \begin{bmatrix} U_{c,1} \Sigma_{c,1} \cdots U_{c,k} \Sigma_{c,k} \cdots U_{c,K} \Sigma_{c,K} \end{bmatrix} \]

\[ = \begin{bmatrix} U_{c,1} \Sigma_{c,1} \cdots U_{c,k} \Sigma_{c,k} \cdots U_{c,K} \Sigma_{c,K} \end{bmatrix} \]

\[ = \begin{bmatrix} V_{c,1}^T \cdots V_{c,k}^T \cdots V_{c,K}^T \end{bmatrix} \]

Composing the parent extended core, $\mathcal{T}_c$: Computation of the extended core associated with the parent whole, $\mathcal{T}_c$, is performed by considering the following derivation
where $\mathbf{T}_{\text{role}} = \mathbf{\Sigma}_{\text{f}}^T \mathbf{\Sigma}_{\text{f}}$. Let $\tilde{T}_c = \tilde{T}_c \times \mathbf{\Sigma}_{c}\mathbf{V}_c^T \times \cdots \times \mathbf{\Sigma}_{c}\mathbf{V}_c^T \times \cdots \times \mathbf{\Sigma}_{c}\mathbf{V}_c^T$. (25)

Overlapping children: This case may be reduced to the non-overlapping case by introducing another level in the hierarchy. Overlapping children are now treated as parents with one non-overlapping child sub-part and child sub-parts that correspond to every possible combination of overlaps that are shared by siblings. The original parent whole representation is computed in terms of the grandchildren representations.

Parent-whole and children-parts with differently filtered data: This is the case when a parent-whole and the children parts contain differently filtered information, as in the case when a parent-whole and the children parts sample information from different layers of a Laplacian pyramid. This case may be reduced to a non-overlapping case by writing the filters as the product between a segmentation filter, $S$, i.e., an identity matrix with limited spatial scope, and general filter that post multiplies the segmentation filter, $\mathbf{H} = \mathbf{F} \cdot S$, and $\mathcal{D} = (\mathcal{D} \times S) \times \mathbf{F}$. The general filters, $\mathbf{F}$, may be applied after the cores are computed.

Computational Cost Analysis: Let an $M$-order data tensor, $\mathcal{D} \in \mathbb{R}^{I_1 \times \cdots I_M}$, where $M = 1 + C$, be recursively subdivided into $K = 2^M$ children of the same order, but with each mode half in size. There are a total of $\log_2 K$ $N + 1$ levels, where $N = \prod_{i=1}^C I_i$. Recursive subdivision results in $S = N \log_2 K$, $N + 1$ segments. The total computational cost is the amortized $M$-mode SVD cost per data tensor segment, $T$, times the number of segments, $O(TN \log_2 K)$). Since siblings at each level can be computed independently, on a distributed system the cost is $O(T \log_2 K)$.

V. CAUSALX EXPERIMENTS

CausalX visual recognition system computes a set of causal explanations based on a counterfactual causal model that takes advantage of the assets of multilinear (tensor) algebra. The $M$-mode Block SVD and the Incremental $M$-mode Block SVD algorithms estimate the model parameters. In the context of face image verification, we compute a compositional hierarchical person representation [42]. Our system is trained on a set of observations that are the result of combinatorially manipulating the scene structure, the viewing and illumination conditions.

We rendered in Maya images of 100 people from 15 different viewpoints with 15 different illuminations. The collection of vectorized images with 10,414 pixels is organized in a data tensor, $\mathcal{D} \in \mathbb{R}^{10,414 \times 15 \times 15 \times 100}$. The counterfactual model is estimated by employing $\mathcal{D}_m$, a hierarchical tensor of part-based Laplacian pyramids. We report encouraging face verification results on two test data sets – the Freiburg, and the Labeled Faces in the Wild (LFW) datasets. We have currently achieved verification rates just shy of 80% on LFW [42], by employing less than one percent (1%) of the total images employed by DeepFace [35]. When data is limited, convolutional neural networks (CNNs) do not convergence or generalize. More importantly, CNNs are predictive rather than causal models.

CONCLUSION

This paper deepens the definition of causality in a multilinear (tensor) framework by addressing the distinctions between intrinsic versus extrinsic causality, and local versus global causality. It proposes a unified multilinear model of wholes and parts that reconceptualizes a data tensor in terms of a hierarchical data tensor. Our hierarchical data tensor is a mathematical instantiation of a tree data structure that enables a single elegant model of wholes and parts and allows for different tree parameterizations for the intrinsic versus extrinsic causal factors. The derived tensor factorization is a hierarchical block multilinear factorization that disentangles the causal structure of data formation. Given computational efficiency considerations, we present an incremental computational alternative that employs the part representations from the lower levels of abstraction to compute the parent whole representations from the higher levels of abstraction in an iterative bottom-up way. This computational approach may be employed to update causal representations in scenarios when data is available incrementally. The resulting object representation is a combinatorial choice of part representations, that renders object recognition robust to occlusion and reduces large training data requirements. We have demonstrated our work in the context of face verification by extending the TensorFaces method with promising results. TensorFaces is a component of CausalX, a counterfactual causal based visual recognition system, and an explainable AI.

ACKNOWLEDGEMENT

The authors are thankful to Ernest Davis for feedback provided during the writing of this document, and to Donald Rubin and Andrew Gelman for helpful discussions.
REFERENCES

[1] E. Acar, E. E. Papalexakis, G. Gürdener, M. A. Rasmussen, A. J. Lawaetz, M. Nilsson, and R. Bro. Structure-revealing data fusion. BMC Bioinformatics, 15(1):239, 2014.

[2] E. Bareinboim and J. Pearl. Causal inference and the data-fusion problem. Proc. of the National Academy of Sciences, 113(27):7345–52, 2016.

[3] P. M. Bentler and S.-Y. Lee. A statistical development of three-mode factor analysis. British J. of Math. and Stat. Psych., 32(1):87–104, 1979.

[4] M. Brand. Incremental singular value decomposition of uncertain data with missing values. In Proc. 7th European Conf. on Computer Vision (ECCV), volume 2350, pages 707–20. Springer, May 2002.

[5] R. Bro. Parafac: Tutorial and applications. In Chemom. Intell. Lab. Syst., Special Issue 2nd Internet Conf. in Chemometrics (INCOC ’95), volume 38, pages 149–171, 1997.

[6] J. D. Carroll and J. J. Chang. Analysis of individual differences in multidimensional scaling via an N-way generalization of ‘Eckart-Young’ decomposition. Psychometrika, 35:283–319, 1970.

[7] W. Chu and Z. Ghahramani. Probabilistic models for incomplete multidimensional arrays. volume 5 of Proceedings of Machine Learning Research, pages 89–96, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA, 16–18 Apr 2009. PMLR.

[8] W. G. Cochran. Observational studies. In T. Bancroft, editor, Statistical Papers in Honor of George W. Snedecor, page 77–90. Iowa State University Press, 1972.

[9] L. de la Lathauwer. Signal Processing Based on Multilinear Algebra. PhD thesis, Katholieke Univ. Leuven, Belgium, 1997.

[10] L. de la Lathauwer. Decompositions of a higher-order tensor in block terms—part i: Definitions and uniqueness. SIAM J. on Matrix Analysis and Applications, 30(3):1033–1066, 2008.

[11] A. Elgammal and C. S. Lee. Separating style and content on a nonlinear manifold. In Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), volume I, pages 478–485, Jun 2004.

[12] A. Gelman and G. Imbens. Why ask why? forward causal inference and reverse causal questions. Tech.report, Nat.Bureau of Econ Research, 2013.

[13] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal. Explaining explanations: An overview of interpretability of machine learning. In 2018 IEEE 5th Inter. Conf. on Data Science and Advanced Analytics (DSAA), pages 80–89. IEEE, 2018.

[14] C. Glymour. Statistics and causal inference: Comment: Statistics and metaphysics. J. of the American Stat. Assoc., 81(396):964–66, Dec 1986.

[15] B. Goodman and S. Flaxman. European union regulations on algorithmic decision-making and a “right to explanation”. AI Magazine, 38(3):50–57, Oct. 2017.

[16] R. Harshman. Foundations of the PARAFAC procedure: Model and conditions for an explanatory factor analysis. Tech. Report Working Papers in Phonetics 16, UCLA, CA, Dec 1970.

[17] P. W. Holland. Statistics and causal inference: Rejoinder. J. of the American Statistical Association, 81(396):968–970, 1986.

[18] E. Hsu, K. Pulli, and J. Popovic. Style translation for human motion. ACM Transactions on Graphics, 24(3):1082–89, 2005.

[19] L. Humberstone. Intrinsically/extrinsic. Synthese, 108:206–267, 1996.

[20] L. O. Vasilescu and D. Terzopoulos. Multilinear projection for appearance-based recognition in the tensor framework. In Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2002), pages 829–832, Champaign, IL, Jun 1992.

[21] M. A. O. Vasilescu and D. Terzopoulos. Multilinear analysis of image ensembles: TensorFaces. In Proc. European Conf. on Computer Vision (ECCV 2002), pages 447–460, Copenhagen, Denmark, May 2002.

[22] M. A. O. Vasilescu and D. Terzopoulos. Multilinear subspace analysis of image ensembles. In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, volume II, pages 93–99, Madison, WI, 2003.

[23] M. A. O. Vasilescu and D. Terzopoulos. TextureSpaces: Multilinear image-based rendering. ACM Transactions on Graphics, 23(3):336–342, Aug 2004. ACM SIGGRAPH 2004 Conf., Los Angeles, CA.

[24] M. A. O. Vasilescu and D. Terzopoulos. Multilinear independent components analysis. In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, volume I, pages 547–553, San Diego, CA, 2005.

[25] D. Vlasic, M. Brand, H. Pfister, and J. Popovic. Face transfer with multilinear models. ACM Transactions on Graphics (TOG), 24(3):426–433, Jul 2005.

[26] H. Wang and N. Ahuja. Facial expression decomposition. In Proc. 9th IEEE Int. Conf. on Computer Vision (ICCV), pages 958–65,3, 2003.

[27] M. Wang, Y. Panagakis, P. Snape, and S. Zafeiriou. Learning the multilinear structure of visual data. In 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pages 6053–6061, Jul 2017.